

DISTRIBUTION ENDPOINT ESTIMATION ASSESSMENT FOR THE USE IN METAHEURISTIC OPTIMIZATION PROCEDURE

Jan Holešovský

Brno University of Technology
 Faculty of Civil Engineering
 Institute of Mathematics and Descriptive Geometry
 Veveří 331/95, 60200 Brno, Czech Republic
 holesovsky.j@fce.vutbr.cz

Abstract: Metaheuristic algorithms are often applied to numerous optimization problems, involving large-scale and mixed-integer instances, specifically. In this contribution we discuss some refinements from the extreme value theory to the lately proposed modification of partition-based random search. The partition-based approach performs iterative random sampling at given feasible subspaces in order to exclude the less favourable regions. The quality of particular regions is evaluated according to the promising index of a region. From statistical perspective, determining the promising index is equivalent to the endpoint estimation of a probability distribution induced by the objective function at the sampling subspace. In the following paper, we give a short review of the recent endpoint estimators derived on the basis of extreme value theory, and compare them by simulations. We discuss also the difficulties in their application and suitability of the estimators for various optimization instances.

Keywords: metaheuristic optimization, endpoint estimation, extreme value, random search, bootstrap, order statistics.

1 Introduction

Metaheuristic algorithms have drawn great attention in the last decades, and nowadays they belong to the most often applied optimization techniques. Among their huge advantages it should be mentioned their adaptability to a wide range of problems and relative suitability in finding good enough (not necessary optimal) solutions. Especially, the metaheuristic algorithms are successfully applied to large-scale optimization problems, possibly involving mixed-integer restrictions. Here the standard techniques exhibit significant limitations. The recently developed and discussed algorithms include, for example, the genetic algorithms [13], particle swarm and ant colony optimization [3], variants of tabu search, and many others [16]. However, such advanced approaches suffer from several shortcomings, primarily the need of quite complex tuning to a specific problem.

In this contribution, we deal with one important class of metaheuristic algorithms known as partition-based random search (PRS). PRS is a probabilistic optimization technique in which the solution space is iteratively divided into smaller segments, whereby any further effort is paid to the most promising regions. A typical PRS algorithm involves the following steps: (i) partitioning of the solution space into smaller subregions; (ii) random sampling within these subregions and determining the corresponding values of an objective function; (iii) evaluation of a so-called promising index of each subregion; (iv) repetition of the foregoing steps with restriction to subregion with the best promising index. The promising index is a suitable measure reflecting the quality of a given subregion, particularly in terms of estimation of the optimal objective function value within the relevant region bounds.

From here on, we concentrate to the maximization problems of the following form

$$\mathbf{y}_{\text{opt}} \in \arg \max_{\mathbf{y} \in \mathcal{Y}} h(\mathbf{y}), \quad (1)$$

where $\mathcal{Y} \subseteq \mathbb{R}^p$ is a non-empty bounded solution space, and the objective function $h : \mathcal{Y} \rightarrow \mathbb{R}$ is continuous and bounded from above. Otherwise there are no other requirements on h or \mathcal{Y} such as convexity etc. For simplicity, we will consider through the paper, that the optimum of the problem (1) is unique.

Let $\mathbf{y}_1, \dots, \mathbf{y}_n$ be a sequence of independent and identically distributed (i.i.d.) random vectors drawn from some probability distribution defined on \mathcal{Y} . Typical choice for \mathbf{y}_i s is that they are drawn from multivariate uniform distribution. Denote X_1, \dots, X_n the corresponding values of the objective function, i.e. $X_i = h(\mathbf{y}_i)$ for $i = 1, \dots, n$. Then X_1, \dots, X_n can be also treated as i.i.d. random variables, say, drawn from a distribution

with cumulative distribution function (c.d.f.) $F(x)$. Denote further $X_{1,n}, \dots, X_{n,n}$ the order statistics obtained from X_1, \dots, X_n such that $X_{1,n} \leq \dots \leq X_{n,n}$.

With respect to the problem (1), the simplest and the most common choice is the evaluation of the promising index as the maximum of the objective function given \mathbf{y}_i s sampled within a region. Thus, the promising index is determined as $X_{n,n}$. As noted in [10], from the statistical perspective is the promising index nothing but an estimator of the endpoint $x^* := \sup\{x : F(x) < 1\}$ of c.d.f. F . The endpoint is the largest value that can a random variable achieve, and hence $x^* = h(\mathbf{y}_{\text{opt}})$. Reliable estimation of x^* plays a crucial role in the PRS algorithm, specifically in order to identify properly the most promising subregion to be restricted to. Hence a suitable estimator \hat{x}^* should be adapted.

The endpoint estimators are usually derived on the basis of the Extreme Value (EV) theory. In [10] the authors proposed, although a minimization problem considered, the probability-weighted moment (PWM) estimator \hat{x}_{PWM}^* based on limiting approximation of the tail through the peaks-over-threshold (POT) method. The PWM estimator will be in detail investigated in Section 2, but let us consider following instance discussed in [10]. Let us split the space \mathcal{Y} into two disjoint subregions $\mathcal{Y}_1, \mathcal{Y}_2$. A small numerical example in [10] shows that the PWM estimator of x^* performs better than the naive estimator $X_{n,n}$ in terms of correct subregion identification (subregion including \mathbf{y}_{opt}). Assume the function to be minimized in (1) is the negated Rastrigin's function $H(\mathbf{y}) := 100 - y_1^2 - y_2^2 + 10 \cos(2\pi y_1) + 10 \cos(2\pi y_2)$, where $\mathbf{y} \in \langle -5, 5 \rangle \times \langle -5, 5 \rangle$. The global optimum $R(\mathbf{y}_{\text{opt}}) = 120$ is attained at the point $\mathbf{y}_{\text{opt}} = (0, 0)$. Moreover, all local maxima are located at points with integer coordinates placed symmetrically around \mathbf{y}_{opt} . In Fig. 1 are plotted the empirical probabilities of identifying the correct subregion. These were obtained from 500 simulations, whereby we the promising index was evaluated as by the sample maxima, so by the PWM estimator applied to the largest values exceeding the 90% sample quantile. Hence, the procedure exactly follows the paper [10]. Two cases of partition were considered: first one with $\mathcal{Y}_1 = \{\mathbf{y} \in \mathcal{Y} : y_1 \leq 0.2\}$, and the second one with $\mathcal{Y}_1 = \{\mathbf{y} \in \mathcal{Y} : y_1 \leq 0.8\}$. Note, because of the positions of local maxima, the set \mathcal{Y}_1 contains the same peaks in both cases.

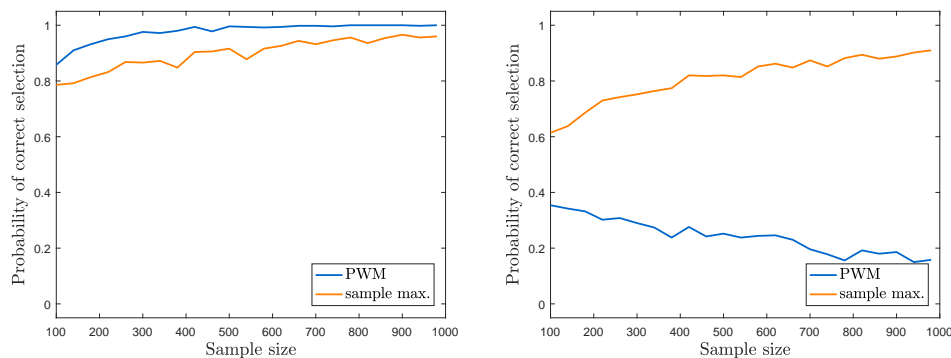


Figure 1: Empirical probabilities of correct subregion identification.

It follows that the PWM endpoint estimation may be very unstable. In the next sections we show that the PWM estimator is not the best choice. Several other estimators have been recently proposed in the literature. We discuss the suitability for the purposes of PRS, and we assess their finite-sample properties by simulations.

2 Endpoint Estimators and Their Properties

Let us consider a continuous distribution with c.d.f. $F(x)$ as above. EV theory is being introduced in terms of the quantile tail functions $U = (1/(1 - F))^\leftarrow$, where f^\leftarrow denotes the left-continuous inverse of f (see [1] for details). The fundamental EV theorem says, that the sample maximum $X_{n,n}$, when properly normalised, converges in distribution to the generalised extreme value (GEV) distribution, if and only if there exists a positive function a such that for $x > 0$ it holds

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} \tag{2}$$

with $\gamma \in \mathbb{R}$. The GEV distribution is a three-parametric family of distributions with shape parameter $\gamma \in \mathbb{R}$, scale parameter $\sigma > 0$, and location parameter $\mu \in \mathbb{R}$. To the condition (2) is referred as to the first-order condition. Of our interest is so-called Weibull class of GEV for $\gamma < 0$, since in this case the (right) endpoint $U(\infty) := x^*$ is finite. Hence, from here on we limit ourselves only to this instance.

All the endpoint estimators discussed below are based on intermediate order statistics $X_{n-k,n}$ where $k = k(n) \rightarrow \infty$ as $n \rightarrow \infty$, and $n/k \rightarrow 0$. This is directly related to POT method applied in [10], whereby

$X_{n-k,n}$ plays a role of a threshold (for POT see e.g. [11]). From relation (2) arise two main types of semi-parametric endpoint estimators. Firstly, for large enough $t = \frac{n}{k}$ and $x \rightarrow \infty$, from (2) it can be derived $U(\infty) \approx U(n/k) - a(n/k)/\gamma$. The quantity $U(n/k)$ is usually estimated by the order statistics $X_{n-k,n}$. The value $a(n/k)$ is asymptotically equivalent to the scale parameter σ . Hence, from here rises the endpoint estimator of the following form [1]

$$\hat{x}^*(k) = X_{n-k,n} - \frac{\hat{\sigma}(k)}{\hat{\gamma}(k)}, \tag{3}$$

where $\hat{\gamma}(k)$ and $\hat{\sigma}(k)$ denote any suitable, k -dependent estimators of γ and σ , respectively.

Another type of the estimators was derived by Fraga Alves and Neves [6]. This is due to expansion of U in (2) in terms of regularly varying functions (see [8] for details). Hereby one obtains

$$\hat{x}_{\text{FAN}}^*(k) = X_{n,n} + \sum_{i=0}^{k-1} a_{ik} (X_{n-k,n} - X_{n-k-i,n}), \tag{4}$$

with $a_{ik} := \log((k+i+1)/(k+i))/\log 2$. Note that the FAN estimator is defined for $k \leq \frac{n+1}{2}$, and the naive endpoint estimator $X_{n,n}$ is just a special case of \hat{x}_{FAN}^* taken $k = 1$.

We consider two estimators of the type (3), the PWM estimator \hat{x}_{PWM}^* and the moment estimator \hat{x}_{MOM}^* . For the first one are the shape and scale parameters estimated by the following

$$\hat{\gamma}_{\text{PWM}} = \frac{\mathcal{M}_0 - 4\mathcal{M}_1}{\mathcal{M}_0 - 2\mathcal{M}_1}, \quad \hat{\sigma}_{\text{PWM}} = \frac{2\mathcal{M}_0\mathcal{M}_1}{\mathcal{M}_0 - 2\mathcal{M}_1},$$

where $\mathcal{M}_i := \frac{1}{k} \sum_{j=1}^k (1 - \frac{j}{k})^i (X_{n-k+j} - X_{n-k,n})$ is the sample PWM of order $(1, r, 0)$ (see [11]). In case of the moment estimator we use

$$\hat{\gamma}_- = 1 - \frac{1}{2} \left(1 - \frac{(M_1)^2}{M_2} \right)^{-1}, \quad \hat{\sigma}_{\text{MOM}} = X_{n-k,n} M_1 (1 - \hat{\gamma}_-),$$

where $M_i := \frac{1}{k} \sum_{j=0}^{k-1} (\log X_{n-j,n} - \log X_{n-k,n})^i$. For PWM there are no restrictions in case of $\gamma < 0$, however the moment estimator $\hat{\gamma}_-$ requires $x^* > 0$. Nevertheless, this additional condition should not provide any serious limitations in applications.

Typical situation in EV analysis is as follows: for small k the estimates are loaded with small bias, but larger variance as only limited number of observations are used for the estimation. Conversely, for k large the variance decreases and the bias component grows. In order to embed the endpoint estimator into the PRS procedure, the estimation precision is of high importance. Thus, the issue to solve is the identification of an optimal number k_0 of order statistics used in (3) or (4). This is commonly based on asymptotic properties of the estimators.

In order to achieve asymptotic distributions of the endpoint estimators, a second-order condition have to be satisfied. Briefly, the second-order condition controls the speed of convergence of the limit in (2). Hereby, it shows up that the convergence can be characterised by a regularly varying function A and a second-order parameter $\rho \leq 0$. The estimation of ρ or the second-order function is quite troublesome, and it is often discussed in the literature [2, 7]. Mostly it requires higher-order conditions to be fulfilled; for further details on second-order condition see e.g. [1]. Hence, in order to avoid the estimation difficulties, advanced techniques are usually applied. This is discussed in Section 3 below.

Then, under some regularity requirements, both \hat{x}_{PWM}^* and \hat{x}_{MOM}^* follow asymptotically a normal distribution with variance and bias depending on γ and both γ and ρ , respectively. Specifically, the bias of \hat{x}_{PWM}^* and \hat{x}_{MOM}^* is managed solely by the second-order framework. On the other hand, the bias of FAN estimator (4) is governed by both the second-order adjustment and the first-order function a in (2). This makes the FAN estimator harder to manage. Moreover, the asymptotic distribution of \hat{x}_{FAN}^* is a mixture of normal and extremal Weibull distribution (see [8], Theorem 2). This leads to different behaviour of the estimator for specific values of γ in terms of both bias and variance. Nevertheless the estimator (4) has several significant advantages. Primarily it does not depend on the estimation of γ , which usually leads to introduction of further uncertainty. Another great property rises from the relation $\hat{x}_{\text{FAN}}^* \geq X_{n,n}$, i.e. the estimator (4) is consistent with the data. This is however guaranteed neither by \hat{x}_{PWM}^* , nor \hat{x}_{MOM}^* . Moreover, the FAN estimator exhibits relatively good stability for wide range of k as discussed in section below.

In order to reduce the first-order bias of \hat{x}_{FAN}^* , a bias-corrected (BC) estimator was proposed [8] of the form

$$\hat{x}_{\text{BC}}^* := \hat{x}_{\text{FAN}}^* - \hat{a} \left(\frac{n}{k} \right) s(\hat{\gamma}), \tag{5}$$

where $\hat{a}(n/k)$ is an estimator of $a(n/k)$, for example $\hat{\sigma}_{\text{MOM}}$ or $\hat{\sigma}_{\text{PWM}}$ can be used as above. The function s is defined as $s(\gamma) := \frac{1}{\gamma} \left(\frac{2^{-\gamma}-1}{\gamma \log 2} + 1 \right)$ and can be obtained from the limit expansion of \hat{x}_{FAN}^* [6]. The estimator \hat{x}_{BC}^* was, however, not closely investigated in [8] as the authors argue with only limited effect of the correction.

3 Simulation Study

In order to assess the behaviour of various endpoint estimators we perform a simulation study taking into account three distributions with c.d.f. limit of sample maxima in the Weibull class (1.–3.), and three benchmark functions (i.–iii.) of various art with unique global maxima. The particular models are listed below. The GEV shape parameters of the latter type can be accessed only through simulation: an extensive, uniformly distributed, random sample of \mathbf{y}_i s was repeatedly drawn on a space \mathcal{Y} , and the value of γ was estimated from the corresponding objective function values. For this purpose was used either the maximum likelihood estimator (suitable for $\gamma > -\frac{1}{2}$, see [1, 17]), or another moment estimator of γ (little different from $\hat{\gamma}_{\text{MOM}}$) [1], whereby the optimal k_0 was chosen via double bootstrap methodology (see e.g. [4, 11]). Hence, we consider the following models:

1. Shifted reverse Burr (RB) distribution with c.d.f. $F(x) = 1 - (1 + (\mu - x)^{-\tau})^{-\lambda}$ for $x \leq \mu = x^*$, where $\tau, \lambda > 0$, and also $\mu > 0$ to ensure $x^* > 0$. The corresponding GEV shape parameter is $\gamma = -1/(\tau\lambda)$.
2. Extremal Weibull distribution (GEV with $\gamma < 0$) with c.d.f. $F(x) = \exp\left\{-\left[1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\}$ for $x \leq \mu - \sigma/\gamma = x^*$.
3. A case of Beta distribution with $F(x) = 1 - (1 - x)^{-1/\gamma}$ for $0 \leq x \leq 1 = x^*$.
 - i. Negative Rosenbrock’s function $R(\mathbf{y}) := 10^5 - 100(y_1^2 - y_2)^2 - (1 - y_1)^2$ on $\mathcal{Y} = \langle -5, 5 \rangle^2$. The global optimum $R(\mathbf{y}_{\text{opt}}) = 10^5$ is attained at $\mathbf{y}_{\text{opt}} = (1, 1)$, $\gamma \approx -2.08$.
 - ii. Negative Rastrigin’s function $H(\mathbf{y}) := 100 - y_1^2 - y_2^2 + 10 \cos(2\pi y_1) + 10 \cos(2\pi y_2)$ on $\mathcal{Y} = \langle -5, 5 \rangle^2$. Here we have $H(\mathbf{y}_{\text{opt}}) = 120$ at $\mathbf{y}_{\text{opt}} = (0, 0)$, $\gamma \approx -0.50$.
 - iii. Gaussian-mixture function $G(\mathbf{y}) := e^{-A(x^2+y^2)} + De^{-B((x-3)^2+(y-3)^2)} + Ee^{-C(x^2+(y-4)^2)}$ on $\mathcal{Y} = \langle -4, 6 \rangle^2$ with $(A, B, C, D, E) = (0.5, 5, 0.1, 0.9, 0.8)$. Global optimum is $G(\mathbf{y}_{\text{opt}}) \approx 1.1964$ at $\mathbf{y}_{\text{opt}} \approx (2.9801, 3.0066)$, $\gamma \approx -0.28$.

The good solutions of $R(\mathbf{y})$ lay inside narrow, parabolic-shaped valley and outside this area the solutions perform poorly. The function $H(\mathbf{y})$ is a composition of periodical function fitted on a paraboloidal support, i.e. there are many local optima with only slightly different values of the objective function. Finally, the function $G(\mathbf{y})$ is a three-peak surface with relative easily distinguishable local and global optima. Moving from these points causes the solution value decrease rapidly to zero. The second best solution takes the value ≈ 1.1706 , hence it is well distinguishable from the global optimum. However, the global-optimum peak is relatively steep, leading sometimes to misspecification by the sample maxima estimates.

In Fig. 2 are plotted the mean endpoint estimates with respect to the choice of k upper order statistics obtained from the models 1.–3. above. The means have been observed from 1,000 independent realizations of size $n = 500$ for various values of γ . Particularly, we considered the following settings of the models: RB with τ variable, $\lambda = 1$, $\mu = 3$; GEV with γ variable, $\sigma = 1$, $\mu = 0$; Beta with γ variable. Note, that the naive estimator $X_{n,n}$ (noted as MAX) is plotted constant, since it coincides to $\hat{x}_{\text{FAN}}^*(1)$. The true value of x^* is indicated by dashed line.

One can see, that the PWM estimator performs the worst in terms of bias. Hereby, when compared to [10], the use of alternative estimators could lead to improvement of the PRS methodology. The moment estimator \hat{x}_{MOM}^* exhibits exactly the bias-variance trade-off discussed above in Section 2. If k is chosen properly, there is chance for zero bias, or better, small bias and concurrently relative small variance, too. The approach for finding k_0 minimizing the mean square error (MSE) proposed in [5] will be discussed below. Nevertheless, both PWM and MOM are generally not data-consistent, i.e. one can obtain clearly unsuitable estimate \hat{x}^* with $X_{n,n} > \hat{x}^*$.

On this account, the naive estimator performs well. However, by its nature, it underestimates the endpoint. The rate of underestimation depends mostly on the tail properties of the distribution and the sample size n . The shape parameter, as indicated by its name, γ influences the length of the right tail of the distribution. While for $\gamma < 0$ the tail is always finite, large negative values of γ indicate the tail is extremely short. On the other hand, γ near zero means there is only a small probability of observation being drawn close enough to the endpoint, such that $X_{n,n}$ could be loaded with small bias. Follow this reasoning in Fig. 2 over the rows, where the bias of $X_{n,n}$ decreases with γ .

The FAN estimator \hat{x}_{FAN}^* overperforms $X_{n,n}$ is the majority of cases, especially for γ close to zero. As noted in [8], the dominant component of the bias comes from the first-order function $a(n/k)$ which, in case γ is close to zero (Fig. 2, first row), exhibits very slow convergence. Still, there is decent possibility to gain smaller bias as that of $X_{n,n}$. As γ decreases, \hat{x}_{FAN}^* exhibits sharper trend, whereby the area for a suitable k selection tightens rapidly. Here we mention the results from [1] and [8], where the authors discuss, that the convergence

of $X_{n,n}$ for $\gamma < -\frac{1}{2}$ is approximately equivalent to the moment related estimators (while \hat{x}_{MOM}^* still shows better stability than \hat{x}_{FAN}^* for $\gamma \leq -0.7$; see last two rows in Fig. 2). Hence, in this case one can benefit from use of the naive estimator in terms of both good precision and simplicity. The authors of [8] discussed shortly the relevance of a bias-corrected FAN estimator \hat{x}_{BC}^* . This estimator overtakes all negatives of regular FAN estimator (i.e. slow convergence for $\gamma \approx 0$, large bias for $\gamma \leq -0.7$). However, in our simulations the estimator \hat{x}_{BC}^* showed only negligible stability improvements for $\gamma = -0.4$. Here we agree with the comments in [8] that the first-order bias-correction in \hat{x}_{BC}^* has very limited effect.

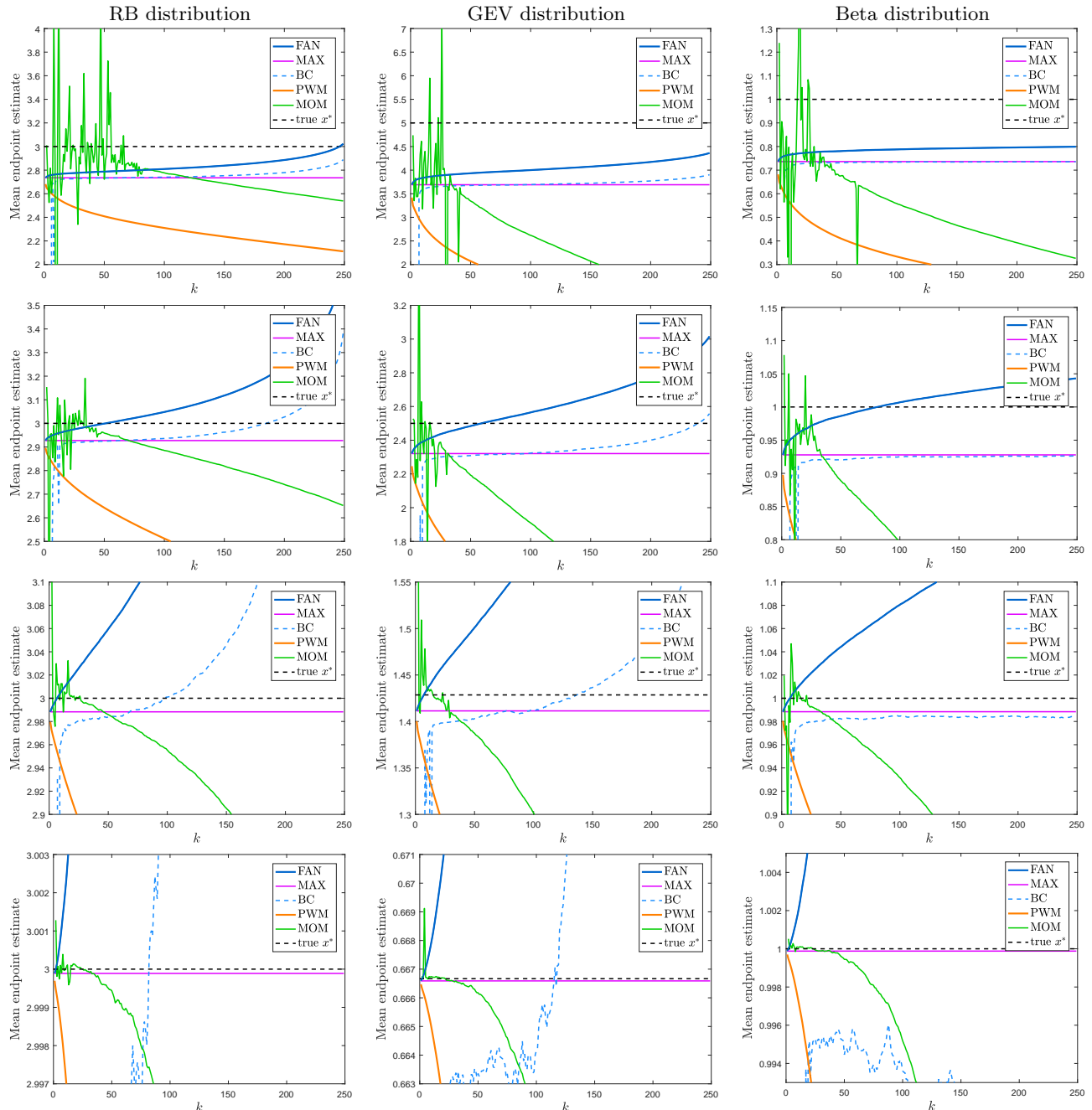


Figure 2: Mean estimates of x^* with respect to the choice of k observed from 1,000 realizations of size $n = 500$. Distributional models: RB (left), GEV (middle), Beta (right). Value γ within rows -0.2, -0.4, -0.7, -1.5.

In Fig. 3 are visualized the mean estimates and MSE of particular endpoint estimators applied to the benchmark functions. As it is well known, for the MSE holds the relation $\text{MSE}(\hat{x}^*) = \text{var}(\hat{x}^*) + \text{bias}(\hat{x}^*)^2$, i.e. minimizing the MSE can be viewed as finding a compromise between variance and bias.

The results in Fig. 3 rather agree with the conclusions made earlier. The large negative value of γ for Rosenbrock's function causes overall estimators to perform badly as the bias grows rapidly with k . Hence, the naive estimator remains the best choice, and this holds also in terms of MSE. In case of the Rastrigin's function both FAN and MOM estimators exhibit good stability for small and intermediate k . However, \hat{x}_{MOM}^* again

suffers from significant variability. This is also visible from the lower plot, where \hat{x}_{MOM}^* shows much larger MSE although in bias it is comparable to FAN. The FAN estimator turns out to be the best choice if k is selected properly. Similar situation was obtained also for the Gaussian-mixture function. Nevertheless, when compared to the simulations made earlier (Fig. 2), the FAN estimate shows a little different behaviour. It converges faster and has a tendency to fluctuate around the true endpoint value.

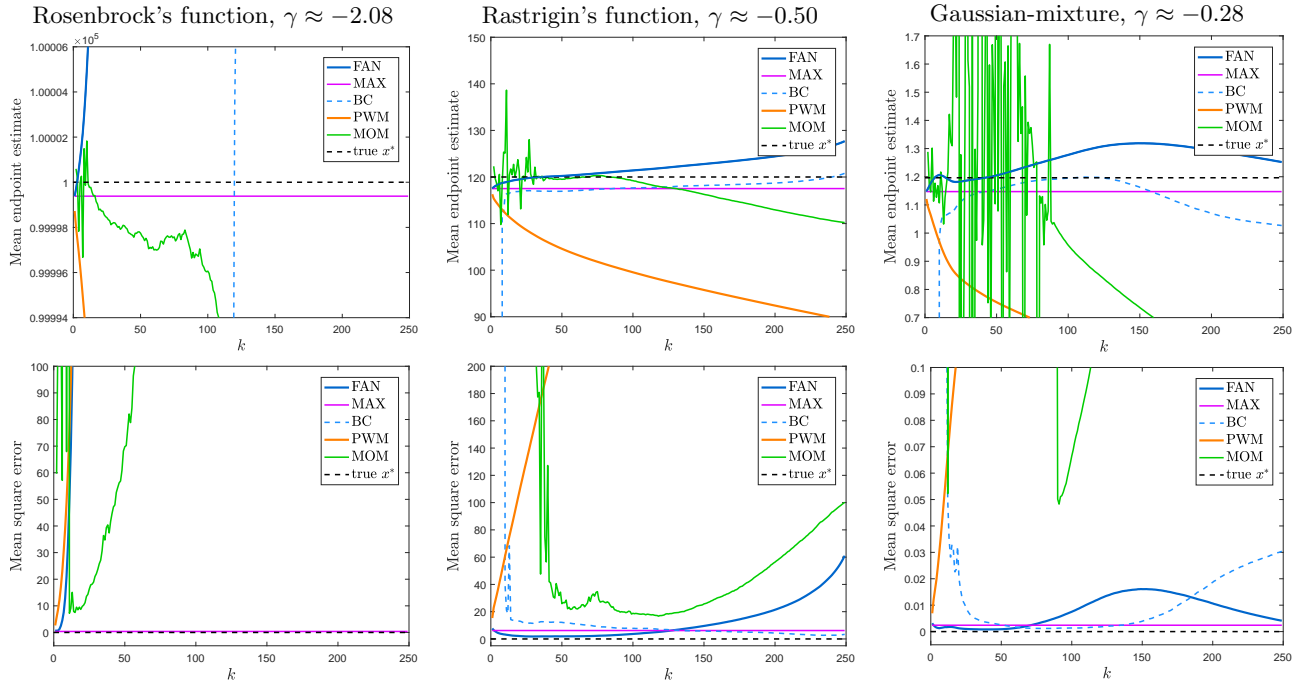


Figure 3: Mean endpoint estimates (upper row) and corresponding MSE (lower row) obtained from 1,000 simulations of size $n = 500$ for the benchmark functions $R(\cdot)$ (left), $H(\cdot)$ (middle), $G(\cdot)$ (right).

Still the main issue to solve is finding an optimal number k_0 of upper order statistics to be embedded into an estimator. A great interest has been paid to this in the area of individual parameter estimation. This is mainly due to the particular importance of the shape parameter γ with respect to the tail probabilities. In practical situations is the optimal k_0 being determined mostly by visual inspection of the parameter dependencies (see e.g. [11] for so-called mean-residual-life plot in POT analysis), or by several rules of thumb (as in [10] where they choose k such that $X_{n-k,n}$ is equivalent to the upper 90% quantile). Lately, attention was paid to adaptive methods for selection of k_0 , follow e.g. the papers [4, 12, 15] or [9]. An exhaustive summary of various approaches is given in [14] including several obsolete ones.

pomoci DB nemusime odhadovat A

In this contribution we consider the double bootstrap methodology of estimating k_0 from [5] developed for the moment estimator \hat{x}_{MOM}^* and $\gamma > -\frac{1}{2}$. The optimal value k_0 is chosen such that the MSE of \hat{x}_{MOM}^* is minimal, i.e. $k_0 \in \arg \min E(\hat{x}_{\text{MOM}}^* - x^*)^2$, where E stands for expected value. Since the value x^* is unknown, it is replaced with an auxiliary estimator \hat{x}_{aux}^* of the form (3) with $\hat{\gamma}(k), \hat{\sigma}(k)$ replaced by third-order moment estimators

$$\hat{\gamma}_3 := 1 - \frac{2}{3} \left(1 - \frac{M_1 M_2}{M_3} \right)^{-1}, \quad \hat{\sigma}_3 := X_{n-k,n} M_1 (1 - \hat{\gamma}_3), \tag{6}$$

with M_i s defined before. This auxiliary endpoint estimator has asymptotic properties similar to \hat{x}_{MOM}^* . The difficulty in assessing k_0 is that the asymptotic MSE depends as on γ , so on the second-order characteristics, i.e. the second-order parameter ρ and the function A . The troublesome estimation of A was shortly outlined in Section 2. The double bootstrap method consists in combination of two different estimates of k_0 . This is done in order to achieve an estimator that is independent on the function A and consists only γ and ρ .

In the double bootstrap methodology a random sample of size $n_1 < n$ is repeatedly, say B times, drawn from the original series X_1, \dots, X_n . Hence we get B new series $X_{1,b}^*, \dots, X_{n_1,b}^*, b = 1, \dots, B$. For each such b -th sample, and $k = 1, \dots, n_1 - 1$, is determined a sequence $q_{n_1,b}^*(k) := (\hat{x}_{\text{MOM}}^*(k) - \hat{x}_{\text{aux}}^*(k))^2$. On this basis is obtained the bootstrap MSE estimator $\frac{1}{B} \sum_{b=1}^B q_{n_1,b}^*(k)$, and minimized with respect to k . Denote the corresponding value $k_0^*(n_1)$. Afterwards, the procedure is repeated with n_1 replaced by $n_2 := \lceil n_1^2/n \rceil$. This yields an optimal value $k_0^*(n_2)$. In [5] it was shown that the value k_0 minimizing MSE of \hat{x}_{MOM}^* is asymptotically

equivalent to \hat{k}_0 determined in the following manner

$$\hat{k}_0(n) := \frac{[k_0^*(n_1)]^2}{k_0^*(n_2)} \cdot \frac{g(\hat{\gamma}_-, \hat{\rho}_{n_1}(k_0^*))}{\bar{g}(\hat{\gamma}_-, \hat{\rho}_{n_1}(k_0^*))}, \tag{7}$$

where $\hat{\rho}_{n_1}(k_0^*) := \frac{\log k_0^*(n_1)}{2 \log(k_0^*(n_1)/n_1)}$ is an estimator of the second-order parameter, $\hat{\gamma}_-$ is the same as in Section 2, and the functions $g(\gamma, \rho)$, $\bar{g}(\gamma, \rho)$ can be found in [5].

In order to assess the performance of the double bootstrap estimation of k_0 we draw 500 independent simulations of size $n = 500$ from the considered benchmark functions. The corresponding values $\hat{x}_{\text{MOM}}^*(k_0)$ with \hat{k}_0 from (7) are compared against the naive estimator. Concurrently, we determine also the estimates observed by \hat{x}_{FAN}^* and \hat{x}_{PWM}^* . The first one is considered with k_0 fixed near the optimum attained in our simulations. Namely, for both the Rastrigin’s and Gaussian-mixture functions we set $k_0 = 50$, for Rosenbrock’s function we put $k_0 = 3$ (just small deviation from $X_{n,n}$ to see the difference). The PWM estimator is evaluated in order to be able to assess the results to the paper [10], and thus we assume $k_0 = \lceil n/10 \rceil$ such that $X_{n-k_0,n}$ is equivalent to the 90% quantile of the data. This is the approach adopted in [10].

Boxplots of the particular estimates are visualized in Fig. 4. Note, as expected, the PWM estimator is significantly biased which makes it unsuitable for any endpoint estimation. This is even enhanced by the heuristic choice of k . On the other hand, the MOM estimator exhibits good performance. The median of MOM is no worse with bias than the naive MAX estimator. Practically, from this perspective it is comparable with the FAN estimator, which was however estimated under the optimal settings. But the MOM estimator suffers from two significant deficits: (i) it is loaded with substantial variance (evident mostly for the Gaussian-mixture function), and (ii) the double bootstrap method for determining \hat{k}_0 is computationally too demanding to be embedded in PRS. It turns out that the only perspective estimator is \hat{x}_{FAN}^* . It has similar small variance as the naive estimator, and can reduce the estimation bias. Nevertheless, there is missing an advisable scheme for the choice of number of order statistics to be used.

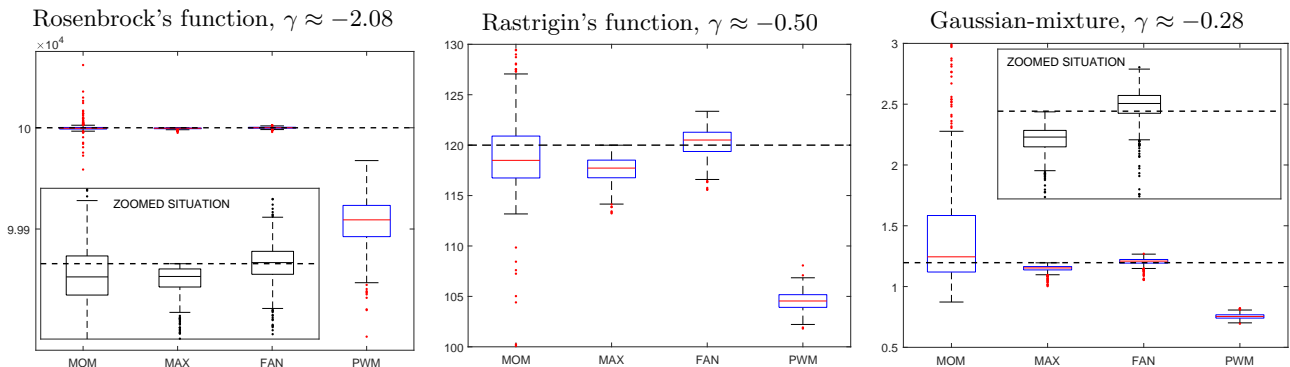


Figure 4: Boxplots of endpoint estimates obtained from 1,000 simulation of size $n = 500$ for the benchmark functions $R(\cdot)$ (left), $H(\cdot)$ (middle), $M(\cdot)$ (right).

4 Conclusion

In this contribution we discussed possible improvements to partition-based random search algorithm. This optimization method has several advantages, primarily it is very flexible with no extraordinary requirements laid on the objective function. Quality of a subregion after partition is given by the promising index. In practical situations, when maximizing an objective function, is such an index constructed simply as a maximum of the objective function attained by a random sample drawn on the region. The authors of [10] proposed evaluation of the promising index on the basis of extreme value estimation of the distribution endpoint. Arguments for that are obvious as the sample maximum has only limited ability to capture the tail properties of a distribution. Other extreme observations within the region are hereby omitted.

Presented simulation study showed that the originally proposed PWM endpoint estimator exhibits very bad finite sample properties. This encompasses especially large bias. Better properties are achieved by the moment estimator. Recently published study [5] provides double bootstrap approach for adaptive selection of tuning parameter k , the number of upper order statistics, based on minimization of the mean square error of the \hat{x}_{MOM}^* . This technique results in good quality estimation, however it can not handle the internal large uncertainty of the moment estimator of γ . Hence, the endpoint estimates are loaded with large variability, too.

Much better results were obtained by the FAN estimator. This estimator is a generalization of the naive estimator $X_{n,n}$. This means it has comparable small variance and it is data-consistent, which is not a feature

governed by neither MOM nor PWM. Moreover, compared to $X_{n,n}$, the FAN estimator can significantly reduce the estimation bias if considered at some suitable level of k . A convenient techniques of proper value of k is however missing, and represents the current challenge for further research.

As noted already in [1], it was proved through simulations, that all the advanced estimators exhibit slow convergence to x^* if the GEV shape parameter γ takes large negative values. In such cases, typically $\gamma < -\frac{1}{2}$, it is more appropriate to use the sample maxima. The rate of convergence to x^* is similar to other estimators, and $X_{n,n}$ is the easiest possible to evaluate.

The statistical methods described above may be also applied within other metaheuristic frameworks, since these lead mainly to solution good enough and not necessary optimal. The endpoint estimation may then serve for quality assessment of such sub-optimal solutions.

Acknowledgement: This contribution was supported by specific research project No. FAST-S-18-5184 at Brno University of Technology and under the project LO1408 *AdMaS UP – Advanced Materials, Structures and Technologies*, supported by Ministry of Education, Youth and Sports of the Czech Republic under the *National Sustainability Programme I*.

References

- [1] de Haan, L., Ferreira, A.: *Extreme Value Theory: An Introduction*. Springer, New-York (2006).
- [2] Deme, E.H., Gardes, L., Girard, S.: On the estimation of the second order parameter for heavy-tailed distribution. *REVSTAT* **11**(3), 277–299 (2013).
- [3] Dorigo, M., Blum, C.: Ant colony optimization: A survey. *Theoretical Computer Science* **344**(2-3), 243–278 (2005).
- [4] Draisma, G., de Haan, L., Peng, L., Pereira, T.T.: A Bootstrap-based Method to Achieve Optimality in Estimating the Extreme-value Index. *Extremes* **2**(4), 367–404 (1999).
- [5] Ferreira, A., de Haan, L., Peng, L.: On optimising the estimation of high quantiles of a probability distribution. *Statistics* **37**(5), 401–434 (2003).
- [6] Fraga Alves, I., Neves, C.: Estimation of the finite right endpoint in the Gumbel domain. *Statistica Sinica* **24**, 1811–1835 (2014).
- [7] Fraga Alves, M.I., Gomes, M.I., de Haan, L.: A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica* **60**(1), 193–213 (2003).
- [8] Fraga Alves, I., Neves, C., Rosário, P.: A general estimator for the right endpoint with an application to supercentenarian women’s records. *Extremes* **20**, 199–237 (2017).
- [9] Gomes, M.I., Oliveira, O.: The bootstrap Methodology in Statistics of Extremes - Choice of the Optimal Sample Fraction. *Extremes* **4**(4), 331–358 (2001).
- [10] Gao, S., Shi, L., Zhang, Z.: A peak-over-threshold search method for global optimization. *Automatica* **89**, 83–91 (2018).
- [11] Holešovský, J., Fusek, M., Blachut, V., Michálek, J.: Comparison of precipitation extremes estimation using parametric and nonparametric methods. *Hydrological Sciences Journal* **61**(3), 2376–2386 (2016).
- [12] Northrop, P.J., Coleman, C.L.: Improved threshold diaostic plots for extreme value analyses. *Extremes* **17**, 289–303 (2014).
- [13] Roupec, J., Popela, P., Hrabec, D., Novotny, J., Olstad, A., Haugen, K.: Hybrid algorithm for network design problem with uncertainty demands. *Lecture Notes in Engineering and Computer Science* **1**, 554–559 (2013).
- [14] Scarrott, C., MacDonald, A.A.: A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT* **10**(1), 33–60 (2012).
- [15] Süveges, M., Davison, A.C.: Model misspecification in peaks over threshold analysis. *Annals of Applied Statistics* **4**(1), 203–221 (2010).
- [16] Talbi, E.-G.: *Metaheuristics: From Design to Implementation*. Wiley, Hoboken (2009).
- [17] Zhou, C.: Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis* **100**, 794–815 (2009).