A horse racing between the block maxima method and the peak–over–threshold approach

Axel Bücher and Chen Zhou

May 6, 2021

Presented by Liujun Chen.

Table of Contents

Domain of Attraction Condition

There exists a constant $\gamma \in \mathbb{R}$ and sequences $a_r > 0$ and $b_r \in \mathbb{R}$ such that

$$\lim_{r\to\infty} F^r(a_r x + b_r) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right) \text{ for all } 1+\gamma x > 0. \quad (1.1)$$

An equivalent representation of the domain of attraction condition is: there exists a positive function $\gamma = \sigma(t)$ such that

$$\lim_{t\to x^*} \frac{1-F(t+\sigma(t)x)}{1-F(t)} = (1+\gamma x)^{-1/\gamma} \text{ for all } 1+\gamma x > 0. \quad (1.2)$$

BM

Let $X_1, \ldots, X_n$ be a sample of observations drawn from F (assume that the observations are independent for the moment).

- Divide the data into $k = [n/r]$ blocks of length r.
- By independence, each block maxima has cdf F^r.
- By (1.1), for large block sizes r, the sample of block maxima can then be regarded as an approximate i.i.d. sample from the three-parametric GEV:

$$G^{GEV}_{\gamma,b,a}(x) := \exp\left\{-(1+\gamma\frac{x-b}{a})^{-1/\gamma}\right\} 1\left(1+\gamma\frac{x-b}{a} > 0\right).$$

- MLE ($\gamma > -1/2$) or PWM ($\gamma < 1/2$).

POT

Equation (1.2) gives rise to the competing peak-over-threshold approach (positive): for sufficiently large t in (1.2), we obtain that, for any $x > 0$,

$$P(X > t + x | X > t) = \frac{P(X > t + x)}{P(x > t)} \approx (1 + \gamma \frac{x}{\sigma})^{-1/\gamma} =: 1 - G^{GP}_{\gamma,\sigma}(x).$$

- In practice, t is typically as the $n - k$th order statistics $X_{n-k,n}$ for some intermediate value k.
- Then, one may regard the sample $X_{n-k+1,n} - X_{n-k,k}, \ldots, X_{n,n} - X_{n-k,n}$ as observations from the two parametric GPD.
- MLE ($\gamma > -1/2$), PWM ($\gamma < 1/2$), Hill ($\gamma > 0$), or Moment ($\gamma \in \mathbb{R}$)

Outline of this paper

The goal of the present paper is an in-depth comparison of the two approaches.

- Efficiency comparison in i.i.d. scenarios.
- BM and POT applied to time series.
- Extensions to multivariate observations and stochastic processes.

Table of Contents

Efficiency Comparison

- The efficiency of BM and POT estimators can be compared in terms of their asymptotic bias and variance. ¡¡¡¡¡¡¡ HEAD =======

  ¿¿¿¿¿¿¿ 289f5da58f43f34b477429583765552fe477880e
- Particularly focus on the estimation of the extreme value index γ.

Efficiency Comparison

- In both the BM and POT approach, a key tuning parameter is $k = k(n)$. (the number of blocks in the BM, the number of upper order statistics in the POT). Theoretically, $k = k(n) \to \infty, k/n \to 0$ as $n \to \infty$.
- The asymptotic variance of the two estimators is of order $1/k$.
- The bias depends on how well the distribution of block maxima or threshold exceedances is approximated by the GEV or GP distribution, respectively.

Two Extreme Cases

Consider two extreme examples first (where the condition $k/n \to 0$ as $n \to \infty$ may in fact be discarded).

- For the Fréchet distribution, the block maximas of size $r = 1$ ($k = n$) are already GEV-distributed. The convergence rate is $1/\sqrt{n}$ and the POT fails to achieve this rate.
- For the POT distribution, $k = n$ largest order statistics can be used for the estimation via POT. The rate of convergence for POT is $1/\sqrt{n}$.

Second Order Condition

- Apart form these two (or similar) extreme cases, the optimal choice of k depends on second order condition quantifying the speed of convergence in the DOA condition.
- They are formulated in terms of the two quantile functions

$$U(x) = \left(\frac{1}{1-F}\right)^{\leftarrow}(x), \;\; V(x) = \left(\frac{1}{-\log F}\right)^{\leftarrow}(x).$$

for the POT- and the BM method, respectively.

DOA condition for U and V

- There exists a positive function a_{POT} such that, for all $x > 0$,

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a_{POT}(t)} = \int_1^x s^{\gamma-1} ds = \frac{x^\gamma - 1}{\gamma} =: h_\gamma(x).$$

- There exists a positive function a_{BM} such that, for all $x > 0$,

$$\lim_{t\to\infty} \frac{V(tx) - V(t)}{a_{BM}(t)} = \int_1^x s^{\gamma-1} ds = \frac{x^\gamma - 1}{\gamma} =: h_\gamma(x).$$

The bias of certain BM- and POT estimators is determined by the speed of convergence in the latter two limit relations, which can be captured by suitable second order conditions.

Second order conditions for U and V

Define $H_{\gamma,\rho}(x) = \int_1^x s^{\gamma-1} \int_1^s u^{\rho-1} duds$.

- Suppose that there exists $\rho_{POT} \leq 0$, a positive function a_{POT} and a positive or negative function A_{POT} with $\lim_{t\to\infty} A_{POT} = 0$, such that for all $x > 0$,

$$\lim_{t\to\infty} \frac{1}{A_{POT}(t)} \left(\frac{U(tx) - U(t)}{a_{POT}(t)} - h_\gamma(x) \right) = H_{\gamma,\rho_{POT}}(x)$$

- Suppose that there exists $\rho_{BM} \leq 0$, a positive function a_{BM} and a positive or negative function A_{BM} with $\lim_{t\to\infty} A_{BM} = 0$, such that for all $x > 0$,

$$\lim_{t\to\infty} \frac{1}{A_{BM}(t)} \left(\frac{V(tx) - V(t)}{a_{BM}(t)} - h_\gamma(x) \right) = H_{\gamma,\rho_{BM}}(x)$$

Remarks about the SOC

- THe functions $|A_{BM}|$ and $|A_{POT}|$ are then necessarily regularly varying with index ρ_{BM} and ρ_{POT} respectively.
- The limit distribution $H_{\gamma,\rho}$ appear specific, see e.g. de Haan and Ferreira (2006).
- If the speed of convergence is faster than any power function, we set $\rho = -\infty$. For example, for GP family, the $\rho_{POT} = -\infty$; for GEV family, $\rho_{BM} = -\infty$.

ρ_{BM} and ρ_{POT}

- It is important to note that ρ_{BM} and ρ_{POT} can be vastly different.
- A general result can be found in Drees, de Haan and Li (2003): If $2tA(t) \to c \in [-\infty, \infty]\backslash(1-\gamma)$, the two coefficients are equal within the range $[-1, 0]$. Otherwise, if $\rho_{BM} < -1$, then $\rho_{POT} = -1$. If $\rho_{POT} < -1$, then $\rho_{BM} = -1$.

ρ_{BM} and ρ_{POT}

Distribution	γ	ρ_{POT}	ρ_{BM}
$\mathrm{GP}(\gamma,\sigma)$	γ	$-\infty$	-1
$\mathrm{Exponential}(\lambda)$	0	$-\infty$	-1
$\mathrm{Uniform}(0,1)$	-1	$-\infty$	-1
Arcsin	-2	-2	-1
$\mathrm{Burr}(\eta,\tau,\lambda)$	$1/(\lambda\tau)$	$-1/\lambda$	$\max(-1/\lambda,-1)$
$t_\nu, \nu \neq 1$	$1/\nu$	$-2/\nu$	$\max(-2/\nu,-1)$
$\mathrm{Cauchy}(=t_1)$	1	-2	-2
$\mathrm{Weibull}(\lambda,\beta), \beta \neq 1$	0	0	0
$\Gamma(\alpha,\beta)$	0	0	0
$\mathrm{Normal}(\mu,\sigma^2)$	0	0	0
$F(x) = \exp(-(1+x^\alpha)^\beta)$	$1/(\alpha\beta)$	$\max(-1/\beta,-1)$	$-1/\beta$
$\mathrm{Fréchet}(\alpha,\sigma)$	$1/\alpha$	-1	$-\infty$
$\mathrm{Reverse\ Weibull}(\beta,\mu,\sigma)$	$-1/\beta$	-1	$-\infty$
$\mathrm{GEV}(\gamma,\mu,\sigma)$	γ	-1	$-\infty$

TABLE 1
Extreme value index and second order parameters for various models.

Asymptotic theory for the estimation of γ

- Solid theoretical studies regarding the POT method have a much longer history. For the sake of theoretical comparability with the BM method, we only deal with the MLE and the PWM.
- Perhaps surprisingly, asymptotic theory for the BM method has hitherto mostly ignored the fact that block maxima are only approximately GEV distributed. ¡¡¡¡¡¡¡ HEAD
- Only recent theoretical studies in Ferreira and de Haan (2015) and Dombry and Ferreira (2019) for the PWM and MLE, respectively, =======
- Only recent theoretical studies in Ferreira and de Haan (2015) and Dombry and Ferreira (2017) for the PWM and MLE, respectively, ¿¿¿¿¿¿¿ 289f5da58f43f34b477429583765552fe477880e take the approximation into account.

Asymptotic theory for the estimation of γ

Under the respective second order conditions for U and V, the asymptotic result can be summarized as

$$\hat{\gamma} \stackrel{d}{\approx} N\left(\gamma + A_m(n/k)b, \frac{1}{k}\sigma^2\right),\ m \in \{BM, POT\}.$$

Here, the asymptotic bias b and the asymptotic variance depend on the specific estimator, ρ_m and γ. In particular, the rate of convergence of the bias $A_m(n/k)$ crucially depends on the second order index ρ_m.

Rate of convergence

We consider the rate of convergence of the root mean squared error.

¡¡¡¡¡¡¡ HEADAssume that $A_m(t) \asymp t^{\rho_m}$ with $\rho_m \in (-\infty, 0)$.The best attainable rate of convergence is achieved when squared bias and variance are of the same order, that is, when

$$A_m^2\left(\frac{n}{k}\right) \asymp \left(\frac{n}{k}\right)^{2\rho_m} \asymp \frac{1}{k}.$$

Solving for k yields $k \asymp n^{-2\rho_m/(1-2\rho_m)}$, which implies =======Assume that $A_m(t) = t^{\rho_m}$ with $\rho_m \in (-\infty, 0)$.The best attainable rate of convergence is achieved when squared bias and variance are of the same order, that is, when

$$A_m^2\left(\frac{n}{k}\right) = \left(\frac{n}{k}\right)^{2\rho_m} = \frac{1}{k}.$$

Solving for k yields $k = n^{-2\rho_m/(1-2\rho_m)}$, which implies ¿¿¿¿¿¿¿ 289f5da58f43f34b477429583765552fe477880e

Rate of convergence of $\hat{\gamma} = n^{-2\rho_m/(1-2\rho_m)}$,

Convergence Rate

2nd Order Parameters	Rate POT	Rate BM	Better rate
$\rho = \rho_{\mathrm{BM}} = \rho_{\mathrm{POT}} \in [-1, 0)$	$n^{\rho/(1-2\rho)}$	$n^{\rho/(1-2\rho)}$	-
$\rho_{\mathrm{BM}} = -1, \rho_{\mathrm{POT}} < -1$	$n^{\rho_{\mathrm{POT}}/(1-2\rho_{\mathrm{POT}})}$	$n^{-1/3}$	POT
$\rho_{\mathrm{POT}} = -1, \rho_{\mathrm{BM}} < -1$	$n^{-1/3}$	$n^{\rho_{\mathrm{BM}}/(1-2\rho_{\mathrm{BM}})}$	BM

- For $\rho_m = -\infty$, the convergence rate is 'faster than $n^{-1/2+\varepsilon}$ for any $\varepsilon > 0$'. And depending on the underlying distribution, in fact could even achieve $n^{-1/2}$.

Asymtotic mean squared error

- If $\rho_{POT} \neq \rho_{BM}$, the approach corresponding to a lower ρ generally yields estimators for γ with a faster attainable rate of convergence than the other approach.
- In this subsection, we consider the case $\rho_{BM} = \rho_{POT}$.
- Hence, the efficiency comparison should be made at the level of asymptotic mean squared error (AMSE) or, more precisely, its two subcomponents: asymptotic bias and asymptotic variance.

Asymtotic mean squared error

- A detailed analysis of the PWM and the ML estimators under the BM and POT approach has ¡¡¡¡¡¡¡ HEAD been carried out in Ferreira and de Haan (2015) and Dombry and Ferreira (2019), for the case $\rho_{BM} = \rho_{POT} = [-1, 0]$ and $\gamma \in (-0.5, 0.5)$. ======= been carried out in Ferreira and de Haan (2015) and Dombry and Ferreira (2017), for the case $\rho_{BM} = \rho_{POT} = [-1, 0]$ and $\gamma \in (-0.5, 0.5)$. ¿¿¿¿¿¿¿ 289f5da58f43f34b477429583765552fe477880e
- When using the same value for k, the BM version of either ML or PWM leads to a lower asymptotic variance compared to the corresponding POT version, for all $\gamma \in (-0.5, 0.5)$.
- Asymptotic bias is smaller for the POT versions of the two estimators, for all $(\gamma, \rho) \in (-0.5, 0.5) \times [-1, 0]$.
- When comparing the optimal AMSE, for the ML, POT is better. For the PWM, BM is preferable for most combinations of (γ, ρ).

Threshold and block length choice

- Both the POT and the BM approach require a practical selection for the intermediate sequence $k = k_n$ in a sample of size n.
- In the POT approach, the choice of k can be interpreted as the choice of the threshold above which the POT approximation s regarded as sufficiently accurate.
- Similarly, in the BM approach, k is related to $r = n/k$, which is the size of the block of which the GEV approximation to the block maximum is regarded as sufficiently accurate.

Threshold and block length choice

- The theoretical conditions that $k \to \infty$ and $k/n \to 0$, as $n \to 0$, are useless in guiding the practical choice.
- Practically, often a plot between the estimates based on various k against the values of k is made for resolving this problem, the so-called "Hill plot".
- The Hill plot can be also be applied to other POT or even BM estimators than just the Hill estimator.
- The ultimate choice is then made by taking a k from the first stable region in the "Hill plot". Nevertheless, the estimators are often rather sensitive to the choice of k.

¡¡¡¡¡¡¡ HEAD

=======

The choice of k for the BM

Unknown Till now.

¿¿¿¿¿¿¿ 289f5da58f43f34b477429583765552fe477880e

Table of Contents

Extremes for Stationary Time Series

- Assume that $(X_t)_{t\in\mathbb{Z}}$ is a strictly stationary univariate time series, and the stationary cdf F satisfies the domain-of-attraction condition. ¡¡¡¡¡¡¡ HEAD ======= ¿¿¿¿¿¿¿

  289f5da58f43f34b477429583765552fe477880e
- It is important to note that the parameters γ, a_r, b_r only depend on the stationary cdf F.

The POT approach for time series

- Recall that the POT approach is based on the sample of large order statistics denoted by $\mathcal{X}_{POT} = \{X_{n-k,n}, \ldots, X_{n,n}\}$.
- Bearing in mind that, under mild extra conditions on the serial dependence (ergodicity, mixing conditions,...), empirical moments are consistent for their theoretical counterparts.
- Thus, we can still use PWM and MLE.
- The asymptotic variance of such estimators will however be different from the i.i.d. case in general.

The POT approach for time series

- Respective theory can be found in Hsing (1991); Resnick and Stărică (1998) for the Hill estimator and in Drees (2000) for a large class of estimators, including PWM and ML.
- Most of the estimators have the same bias as in the i.i.d. case, whereas their asymptotic variances depend on the serial dependence structure and are usually higher than those obtained in the i.i.d. case.
- Since the asymptotic bias shares the same explicit form, bias correction can also be performed in the same way as in the i.i.d. case; see, e.g., de Haan, Mercadier and Zhou (2016).

The BM approach for time series

- Recall that the BM approach is based on the sample of block maxima $\mathcal{X}_{BM} = \{M_{1,r}, \dots, M_{k,r}\}$, where $M_{j,r}$ denotes the maximum within the jth disjoint block of observations of size r.
- The main motivation lead us to consider this sample as approximately GEV-distributed was the relation

$$P\left(M_{1,r} \leq a_r x + b_r\right) = F^r(a_r x + b_r) \approx G_{\gamma,0,1}^{GEV}(x),$$

for large r.
- The first equality is not true for time series, whence more sophisticated arguments must be found for the BM method to work for time series.

The BM approach for time series

- If F satisfies the DOA condition, and some mixing condition on the serial dependence are met, then there exists a constant $\theta \in [0,1]$ such that

$$P\left(M_{1,r} \leq a_r x + b_r\right) \to \left(G^{GEV}_{\gamma,0,1}(x)\right)^{\theta}.$$

- The constant θ is called the extremal index and can be interpreted as capturing the tendency of the time series that extremal observations occur in clusters.

Extremal Index

If $\theta < 0$, then letting

$$\tilde{a}_r = a_r \theta^\gamma, \tilde{b}_r = b_r - a_r \frac{1-\theta^\gamma}{\gamma},$$

we immediately obtain that

$$P\left(M_{1,r} \leq \tilde{a}_r x + \tilde{b}_r\right) \to G^{GEV}_{\gamma,0,1}(x).$$

Estimation of the extremal index

- 1) BM-like estimators based on “blocking” techniques (Northrop, 2015; Berghaus and Bücher, 2017)
- 2) POT-like estimators that rely on threshold exceedances (Ferro and Segers, 2003; Süveges, 2007)
- 3) estimators that use both principles simultaneously (Hsing, 1993; Robert, 2009; Robert, Segers and Ferro, 2009)
- 4) estimators which, next to choosing a threshold sequence, require the choice of a run-length parameter (Smith and Weissman, 1994; Weissman and Novak, 1998).

BM for time series

- Since the distance between the time points at which the maxima within two successive blocks are attained is likely to be quite large, the sample $\mathcal{X}_{BM}$ can be regarded as approximately independent.
- As a matter of fact, the literature on statistical theory for the BM method is mostly based on the assumption that $\mathcal{X}_{BM}$ is a genuine i.i.d. sample from the GEV-family.
- Two approximation errors are thereby completely ignored: the cdf is only approximately GEV, and the sample is only approximately independent.

BM for time series

- Solid theoretical results taking these errors into account are rare: Bücher and Segers (2018b) treat the ML-estimator in the heavy-tailed case ($\gamma > 0$).
- The main conclusions are: the sample can safely be regarded as independent, but a bias term may appear which, similar as in Section 2, depends on the speed of convergence.(SOC)
- Bücher and Segers (2018a) improve upon that estimator by using sliding blocks instead of disjoint blocks.
- The asymptotic variance of the estimator decreases, while the bias stays the same. Moreover, the resulting 'Hill-Plots' are much smoother, guiding a simpler choice for the block length parameter.

Comparisons

- Due to the lack of a general theoretical result on the BM method, a theoretical comparison on which method is more efficient seems out of reach for the moment.
- However, some insight into the merits and pitfalls of two approaches can be gained by considering the problem of estimating high quantiles and return levels.

Estimating high quantiles

For the BM,

$$F^{\leftarrow}(1-p) \approx b_r + a_r \frac{\{-r\log(1-p)\}^{-\gamma}-1}{\gamma} \approx b_r + a_r \frac{(rp)^{-\gamma}-1}{\gamma}.$$

For the POT,

$$F^{\leftarrow}(1-p) \approx t + \sigma(t) \frac{(\frac{p}{1-F(t)})^{-\gamma}-1}{\gamma}.$$

As a consequence, based on the plug-in principle, the POT method immediately yields estimators for high quantiles On the other hand, the BM method cannot be used straight forwardly. It requires the estimation for θ.

Estimating return levels

Let $F_r(x) = P(M_{1,r} \leq x)$. For $T \geq 1$, the T -return level of the sequence of block maxima is defined as the $1 - 1/T$ quantile of F_r, that is

$$RL(T, r) = F_r^{\leftarrow}(1 - 1/T).$$

We have that

$$RL(T, r) \approx \tilde{b}_r + \tilde{a}_r \frac{(r/T)^{-\gamma} - 1}{\gamma}.$$

Following the discussion in the previous section, it is now the BM method which yields simpler estimators that do not require additional estimation of the extremal index.

Table of Contents

Multivariate Extremes

suppose that there exists a non-degenerate cdf G and sequences $(a_{r,j})_{r\in\mathbb{N}}$, $(a_{r,j})_{r\in\mathbb{N}}$, $j=1,\ldots,d$ with $a_{r,j}>0$ such that

$$\lim_{r\to\infty} P\left(\frac{\max_{i=1}^r X_{i,1}-b_{r,1}}{a_{r,1}}\le x_1,\ldots,\frac{\max_{i=1}^r X_{i,d}-b_{r,d}}{a_{r,d}}\le x_d\right)$$
$$= G(x_1,\ldots,x_d).$$

The dependence between the coordinates of G can be described in various equivalent ways:

- by the stable tail dependence function L (Huang, 1992),
- by the exponent measure μ (Balkema and Resnick, 1977),
- by the Pickands dependence function A (Pickands, 1981),
- by the tail copula Λ (Schmidt and Stadtmüller, 2006),
- by the spectral mea- sure Φ (de Haan and Resnick, 1977),
- by the madogram ν (Naveau et al., 2009),

Multivariate Extremes

- Often, estimation of the marginal parameters and of the dependence structure is treated successively.
- Standard errors for estimators of the dependence structure may then be influenced by standard errors for the marginal estimation, (a point which is often ignored in the literature on statistics for multivariate extremes.)

The POT method in the multivariate case

Based on observations

$$\mathcal{X}_{POT} = \left\{ \mathrm{X}_i | \mathrm{rank}(X_{i,j} \text{ among } X_{1,j,\dots,x_{n,j}}) \geq n - k \text{ for some } j = 1, \dots, d \right\}$$

e.g. The estimation for the L function.
Bias correction for $\hat{L}$ function have been proposed in in Fougères et al. (2015).

The BM method in the multivariate case

- Let r denote a block size, and $k = [n/r]$ the number of blocks.
- For $l = 1, \ldots, k$, let $\mathrm{M}_{l,r} = (M_{l,1,r}, \ldots, M_{l,d,r})'$ denote the vector of componentwise block maxima.
- Any estimator defined in terms of the sample $\mathcal{X}_{BM} = \{\mathrm{M}_{1,r}, \ldots, \mathrm{M}_{k,r}\}$ is called an estimator based on the BM approach.
- Just as for the univariate BM method, asymptotic theory is usually formulated under the assumption that $M_{1,r}, \ldots, M_{k,r}$ are iid sample from the limiting distribution G.
- Moreover, estimation of the marginal parameters is often disentangled from estimation of the dependence structure, with theory for the latter either developed under the assumption that marginals are completely known.

Conclusion

There is no winner.