TAIL INFERENCE USING EXTREME U-STATISTICS

Jochem Oorschot, Johan Segers, Chen Zhou

LIDAM Discussion Paper ISBA
2022 / 14

UCLouvain

LIDAM
ISBA | Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA
Voie du Roman Pays 20 - L1.04.01
B-1348 Louvain-la-Neuve
Email : lidam-library@uclouvain.be
https://uclouvain.be/en/research-institutes/lidam/isba/publication.html

Electronic Journal of Statistics
Vol. 0 (0000) 1
ISSN: 1935-7524
DOI: 10.1214/154957804100000000

Tail inference using extreme U-statistics

Jochem Oorschot

Econometric Institute, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands. e-mail: oorschot@ese.eur.nl

Johan Segers

LIDAM/ISBA, UCLouvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. e-mail: johan.segers@uclouvain.be

Chen Zhou

Econometric Institute, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands. e-mail: zhou@ese.eur.nl

Abstract: Extreme U-statistics arise when the kernel of a U-statistic has a high degree but depends only on its arguments through a small number of top order statistics. As the kernel degree of the U-statistic grows to infinity with the sample size, estimators built out of such statistics form an intermediate family in between those constructed in the block maxima and peaks-over-threshold frameworks in extreme value analysis. The asymptotic normality of extreme U-statistics based on location-scale invariant kernels is established. Although the asymptotic variance corresponds with the one of the Hájek projection, the proof goes beyond considering the first term in Hoeffding's variance decomposition; instead, a growing number of terms needs to be incorporated in the proof.

To show the usefulness of extreme U-statistics, we propose a kernel depending on the three highest order statistics leading to an unbiased estimator of the shape parameter of the generalized Pareto distribution. When applied to samples in the max-domain of attraction of an extreme value distribution, the extreme U-statistic based on this kernel produces a location-scale invariant estimator of the extreme value index which is asymptotically normal and whose finite-sample performance is competitive with that of the pseudo-maximum likelihood estimator.

Keywords and phrases: U-statistic, Generalized Pareto distribution, Hájek projection, Extreme value index.

Contents

1. Introduction

Let $X_1, \dots, X_n$ be independent and identically distributed random variables drawn from a common distribution F. Consider the generalized average

$$U_n^m = \binom{n}{m}^{-1} \sum_{I \subset [n], |I|=m} K_m(X_I), \tag{1.1}$$

where $K_m : \mathbb{R}^m \to \mathbb{R}$ with $1 \leq m \leq n$ is a permutation invariant function called kernel, $[n] := \{1, \dots, n\}$ and X_I is the vector $(X_i)_{i \in I}$ indexed by the subset I. The generalized average U_n^m is called a U-statistic and is an estimator of its corresponding "kernel parameter" θ_m,

$$\theta_m = \mathbb{E}[U_n^m] = \mathbb{E}[K_m(X_1, \dots, X_m)].$$

By efficiently exploiting the information in the sample, U-statistics corresponding to kernels that are not linked to a particular parametric model can attain uniformly minimum variance among all unbiased estimators of θ_m, see [9] for the earliest theoretical development and [12], [18, Chapter 5] or [19, Chapter 12] for an overview of U-statistics.

In this work we adapt the theory of U-statistics to the setting of extreme value analysis, which is concerned with the tail of a distribution function. If the block size m is held fixed, as in the classic setup of U-statistics, there are many blocks containing observations that cannot be considered as part of the tail of the distribution. Therefore, we require that the block size m tends to infinity as $n \to \infty$ and we let the kernel function K_m depend only on the top-q observations in a block of m observations, i.e., there is $K : \mathbb{R}^q \to \mathbb{R}$ satisfying

$$K_m(X_1, \dots, X_m) = K(X_{m:m}, \dots, X_{m-q+1:m}), \tag{1.2}$$

where $X_{m:m} \geq X_{m-1:m} \geq \dots \geq X_{1:m}$ denote the order statistics of $X_1, \dots, X_m$. Note that while m grows, the number of upper order statistics q and the kernel K on $\mathbb{R}^q$ remain fixed. U-statistics corresponding to kernel functions as in (1.2) where $m := m(n) \to \infty$ and $m = \mathrm{o}(n)$ as $n \to \infty$ are called extreme U-statistics.

Following usual practice in extreme value analysis, we assume that the observations are drawn from a common distribution function F in the domain of attraction (DoA) of a Generalized Extreme Value (GEV) distribution with parameter $\gamma \in \mathbb{R}$. For such F, there exist sequences a_n and b_n such that

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma}) \tag{1.3}$$

for $x \in \mathbb{R}$ such that $1 + \gamma x > 0$; notation $F \in \mathcal{D}(G_\gamma)$. The parameter γ is called the extreme value index and governs the tail behaviour of F. We will show that extreme U-statistics can be valid estimators of $\lim_{m\to\infty} \theta_m = \mu(\gamma)$ for some function μ. If the function μ is invertible, we can use the extreme U-statistic to construct estimators of γ.

Extreme U-statistics corresponding to location-scale invariant kernels were first explored in [16]. It was shown there that under suitable conditions the extreme U-statistic is consistent for $\mu(\gamma)$. In this work, we study the asymptotic distribution of extreme U-statistics and we will show that, like classic U-statistics [10], they are asymptotically normally distributed. Closely related to our work is the all-block-maxima method in [14], at the heart of which lie extreme U-statistics with kernels K of degree $q = 1$. Considering higher degrees considerably enriches the possibilities. In particular, we will propose a location-scale invariant kernel K of degree $q = 3$ such that $\mu(\gamma) = \gamma$, leading to a new estimator for the extreme value index with good finite-sample properties.

A technical challenge for extreme-U statistics is that their asymptotic distribution cannot be obtained in the same way as for classic U-statistics. A new approach is needed, which constitutes our main theoretical contribution. Recall that, classically, one truncates the variance of the U-statistic to the first term of the Hoeffding variance decomposition [10, equation (5.12)] and exploits that the U-statistic's Hájek projection has the same variance. In contrast, we need to take into account more than the first term of the Hoeffding decomposition to deal with the asymptotic variance of the extreme U-statistic. In Proposition 2.3 we show that the asymptotic variance of the extreme U-statistic and the variance of the Hájek projection are asymptotically equal by using the first ℓ_n terms of the Hoeffding decomposition for an appropriately chosen sequence $\ell_n \to \infty$ as $n \to \infty$. Note that the ℓ-th term in the Hoeffding decomposition is related to the covariance between the values that result from applying the kernel function to two different blocks with ℓ common elements. We show that such a covariance is dominated by scenarios where a single one of the common elements enters the top q of both blocks. For details, see the proof of Proposition 2.3 in Section A and Section A.1 in particular.

For a distribution function F that is in the DoA of a GEV distribution and for which the kernel expectation θ_m converges to $\mu(\gamma)$, Theorem 3.1 provides the asymptotic normality of extreme U-statistics under additional mild conditions. In Corollary 3.2, we fix $q = 3$, the minimum number of top order statistics required to obtain a location-scale invariant kernel, and we provide easily verifiable kernel integrability conditions guaranteeing that Theorem 3.1 applies.

In statistical applications, estimators derived from extreme U-statistics have the potential to be more efficient than estimators based on other block techniques in extreme value analysis. To see that, note that for classic U-statistics, irrespective of the choice of the kernel K_m, the variance of U_n^m is bounded by

$$\operatorname{var} U_n^m \leq \frac{1}{k} \operatorname{var} K_m(X_1, \ldots, X_m) \tag{1.4}$$

where,for convenience, $k = n/m$ is assumed to be an integer. The disjoint block estimator

$$B_n^m = \frac{1}{k} \sum_{i=1}^{k} K_m(X_{(i-1)m+1}, \ldots, X_{im}), \tag{1.5}$$

that partitions the sample into k consecutive blocks of size m satisfies $\mathbb{E}[B_n^m] = \theta_m$ too, but

$$\operatorname{var} B_n^m = \frac{1}{k} \operatorname{var} K_m(X_1, \ldots, X_m),$$

the upper bound in (1.4). Therefore U_n^m is at least an improvement over the disjoint block estimator. In the context of extreme value analysis, block techniques are often used to extract large observations, for example, in the classical block maxima approach, which has received renewed interest in, for instance, [5, 7]. In addition, recent studies have proposed estimators based on the sliding block technique. Effectively, such estimators increase the number of blocks used in estimation and also have a lower asymptotic variance than B_n^m [1, 15]. Compared to sliding block techniques, extreme U-statistics use even more blocks by exhausting the space of all possible blocks.

As an application of Corollary 3.2, we consider in Section 4 a location-scale invariant kernel $K_* : \mathbb{R}^3 \to \mathbb{R}$ that is unbiased for $\gamma \in \mathbb{R}$ if the underlying data are Generalized Pareto (GP) distributed with parameter γ. Here a random variable Z is GP(γ) distributed if $\mathbb{P}(Z \leq x) = 1 - (1 + \gamma x)^{-1/\gamma}$ for $1 + \gamma x > 0$. The corresponding extreme U-statistic $U_{n,*}^m$, as estimator of γ, can be applied under very weak conditions, e.g. when the data are sampled from a distribution with a bounded density function and in the DoA of a GEV distribution G_γ. In a simulation study, we find that $U_{n,*}^m$ is competitive with the location-scale invariant GP Maximum Likelihood (ML) estimator [3] when the observations are in fact GP distributed and that $U_{n,*}^m$ generally outperforms the GP ML estimator for other distributions in the DoA of a GEV distribution.

Finally, we remark that extreme U-statistics can also be written as a weighted sum of the kernel K evaluated in combinations of order statistics of the sample:

$$U_n^m = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < i_2 < \ldots < i_q \leq n-m+q} \binom{n - i_q}{m - q} K(X_{n-i_1+1:n}, \ldots, X_{n-i_q+1:n}).$$

Such representations are particularly useful for calculating U_n^m in practice, see for example Lemma 4.4. Moreover, the weights $p_{i_q} := \binom{n}{m}^{-1} \binom{n-i_q}{m-q}$ have already appeared in likelihood equation (2.3) in [14] for $q = 1$. That expression can be seen as a particular instance of an extreme U-statistic with $q = 1$.

In Section 2, we develop the asymptotic theory of extreme U-statistics based on samples of the generalized Pareto distribution. The extension to samples from a distribution in the DoA of some GEV distribution is treated in Section 3. A kernel leading to a novel estimator of the extreme value index is constructed in Section 4, the finite-sample performance of which is reported in Section 5. After a brief discussion in Section 6, the proofs of the results are given in a series of appendices.

2. Asymptotic theory of extreme U-statistics

Under conditions typically invoked in extreme value analysis, the extreme U-statistic U_n^m is consistent for a limit quantity $\mu(\gamma)$ with the usual speed of convergence $U_n^m - \mu(\gamma) = O_p(k^{-1/2})$, where $k = n/m$. Moreover, the asymptotic distribution of U_n^m, for a random sample drawn from a distribution function $F \in \mathcal{D}(G_\gamma)$, satisfies

$$\sqrt{k}\,(U_n^m - \mu(\gamma)) \xrightarrow{d} N\left(\lambda B_K(\rho,\gamma), \sigma_K^2(\gamma)\right), \tag{2.1}$$

for a sufficiently integrable location-scale invariant kernel K, where $\xrightarrow{d}$ indicates convergence in distribution, λ and $B_K(\rho,\gamma)$ are deterministic constants to be specified.

The result (and its proof) is naturally subdivided into two parts: first, in this section, we show that relation (2.1) holds for the U-statistic $U_{n,Z}^m$ associated with the GP(γ) distributed sample $Z_1,\ldots,Z_n$, with $B_K(\rho,\gamma) = 0$. Second, in Section 3, we show that the weak limit of $\sqrt{k}\,(U_n^m - \mu(\gamma))$ for a random sample drawn from a common distribution F in the DoA of a GEV distribution with shape parameter γ is the same as the weak limit of $\sqrt{k}\left(U_{n,Z}^m - \mu(\gamma)\right)$ plus the asymptotic bias $\lambda B_K(\rho,\gamma)$.

To appreciate the special role of the GP(γ) distribution, we introduce some distributional representations in Section 2.1. We continue in Section 2.2 with the asymptotic normality of the U-statistic $U_{n,Z}^m$ based on a GP(γ) sample stated in Theorem 2.5.

2.1. Preliminary distributional representations

Let $\stackrel{d}{=}$ denote equality in distribution. For a distribution function F, define the tail quantile function as

$$U(x) := \inf\{x \in \mathbb{R} : (1 - F(x))^{-1} \geq y\}, \qquad y > 1. \tag{2.2}$$

The tail quantile function of the standard GP distribution with parameter $\gamma \in \mathbb{R}$ is

$$h_\gamma(y) = \int_1^y u^{\gamma-1}\mathrm{d}u = \begin{cases} (y^\gamma - 1)/\gamma & \text{if } \gamma \neq 0, \\ \ln y & \text{if } \gamma = 0. \end{cases} \tag{2.3}$$

A random variable Y is said to have a standard Pareto distribution if $\mathbb{P}(Y \leq y) = 1 - 1/y$ for $y \geq 1$. The function h_γ has the property that if Y follows a standard Pareto distribution, the variable $Z = h_\gamma(Y)$ follows a $\mathrm{GP}(\gamma)$ distribution. Moreover, the function h_γ satisfies the functional equation

$$h_\gamma(xy) = h_\gamma(x) + x^\gamma\, h_\gamma(y), \qquad x, y > 0.$$

Let $Y_1, Y_1^*, Y_2, Y_2^*, \ldots$ be independent standard Pareto distributed random variables. For integer $1 \leq q \leq m$, Rényi's representation for exponential order statistics implies that the vector of top-q order statistics of $Y_1^*, \ldots, Y_m^*$ admits the representation

$$(Y^*_{m-i+1:m})_{i=1}^q \stackrel{d}{=} (Y^*_{m-q+1:m} Y_{q-i:q-1})_{i=1}^q \tag{2.4}$$

where $Y_{0:q-1} = 1$. Then, we can represent the vector of top-q order statistics of an independent random sample $Z_1^*, \ldots, Z_n^*$ from the $\mathrm{GP}(\gamma)$ distribution as

$$\begin{aligned}(Z^*_{m-i+1:m})_{i=1}^q &\stackrel{d}{=} (h_\gamma(Y^*_{m-i+1:m}))_{i=1}^q \\ &\stackrel{d}{=} (h_\gamma(Y^*_{m-q+1:m} Y_{q-i:q-1}))_{i=1}^q \\ &= \big(h_\gamma(Y^*_{m-q+1:m}) + (Y^*_{m-q+1:m})^\gamma h_\gamma(Y_{q-i:q-1})\big)_{i=1}^q .\end{aligned}$$

We have thereby proved the following result.

Lemma 2.1. *Consider the integers $1 \leq q \leq m$ and the scalar $\gamma \in \mathbb{R}$. The vector of top-q order statistics of an independent $\mathrm{GP}(\gamma)$ random sample $Z_1^*, \ldots, Z_m^*$ admits the representation*

$$(Z^*_{m-i+1:m})_{i=1}^q \stackrel{d}{=} \big(Z^*_{m-q+1:m} + (1 + \gamma Z^*_{m-q+1:m}) Z_{q-i:q-1}\big)_{i=1}^q$$

where $Z_1, \ldots, Z_{q-1}$ is another independent standard $\mathrm{GP}(\gamma)$ random sample, independent from $Z_1^*, \ldots, Z_m^*$, and where $Z_{0:q-1} = 0$.

Recall that the kernel $K_m : \mathbb{R}^m \to \mathbb{R}$ of the extreme U-statistic is such that, for all integer $m \geq q$ and all $(x_1, \ldots, x_m) \in \mathbb{R}^m$,

$$K_m(x_1, \ldots, x_m) = K(x_{m:m}, \ldots, x_{m-q+1:m}),$$

for some (symmetric) function $K : \mathbb{R}^q \to \mathbb{R}$. If the kernel is also location-scale invariant, i.e., if the kernel satisfies

$$K_m(ax_{m:m} + b, \ldots, ax_{m-q+1:m} + b) = K(x_{m:m}, \ldots, x_{m-q+1:m}).$$

for all $a > 0$ and $b \in \mathbb{R}$, the representation derived in Lemma 2.1 has a useful consequence.

Lemma 2.2. *Consider integers $3 \leq q \leq m$ and $\gamma \in \mathbb{R}$. If the kernel $K : \mathbb{R}^q \to \mathbb{R}$ is location-scale invariant and if $Z_1^*, \ldots, Z_m^*$ are independent standard $\mathrm{GP}(\gamma)$ random variables, then*

$$\big(Z^*_{m-q+1:m}, K_m(Z_1^*, \ldots, Z_m^*)\big) \stackrel{d}{=} \big(Z^*_{m-q+1:m}, K\big((Z_{q-i:q-1})_{i=1}^q\big)\big)$$

with $Z_1, \ldots, Z_{q-1}$ as in Lemma 2.1. In particular, $K_m(Z_1^*, \ldots, Z_m^*)$ is independent of $Z^*_{m-q+1:m}$ and its distribution does not depend on m.

It follows from Lemma 2.2 that

$$\theta := \theta_m = \mathbb{E}\left[K_m\left(Z_1^*, \ldots, Z_m^*\right)\right] = \mathbb{E}\big[K\big((Z_{q-j:q-1})_{j=1}^q\big)\big] = \mu(\gamma), \tag{2.5}$$

which does not depend on m, provided $m \geq q$. Hence, if the observations are drawn from the GP(γ) distribution, the bias term $B_K(\rho, \gamma)$ is expected to be zero in relation (2.1).

2.2. *Random sample drawn from a GP distribution*

In this section we will handle the extreme U-statistic

$$U_{n,Z}^m = \binom{n}{m}^{-1} \sum_{I \subset [n], |I|=m} K_m(Z_I), \tag{2.6}$$

for independent random variables $Z_1, \ldots, Z_n$ drawn from the GP(γ) distribution. We start with a sketch of the derivation of the asymptotic distribution of a U-statistic in the classical case where m is fixed [18, 19], which will provide guidance in the case where $m \to \infty$. It will turn out that the same proof technique can be applied to the case where $m \to \infty$ but $m^2 = \mathrm{o}(n)$ as $n \to \infty$, leading to the asymptotic distribution of the extreme U-statistics. By contrast, if $m^2 = \mathrm{o}(n)$ does not hold, a novel approach is required: in the Hoeffding variance decomposition, a growing number of terms needs to be taken into account.

Classical theory for fixed m. By Hoeffding's decomposition,

$$\operatorname{var} U_{n,Z}^m = \sum_{\ell=1}^m p_{n,m}(\ell)\zeta_{m,\ell} \quad \text{with} \quad \begin{cases} p_{n,m}(\ell) = \binom{n}{m}^{-1}\binom{m}{\ell}\binom{n-m}{m-\ell}, \\ \zeta_{m,\ell} = \operatorname{cov}\big(K_m(X_S), K_m(X_{S'})\big), \end{cases} \tag{2.7}$$

where S and S' are sets of positive integers with $|S| = |S'| = m$ and $|S \cap S'| = \ell$. For fixed m, we get from (2.7) the asymptotic relation

$$\frac{p_{n,m}(\ell+1)}{p_{n,m}(\ell)} = \mathrm{O}\left(n^{-1}\right), \qquad \text{as } n \to \infty. \tag{2.8}$$

Provided $\zeta_{m,1} \neq 0$, it follows that the first term of Hoeffding's decomposition, $p_{n,m}(1)\zeta_{m,1}$, asymptotically dominates the value of the sum $\sum_{\ell=1}^m p_{n,m}(\ell)\zeta_{m,\ell}$. For $p_{n,m}(1)$, we find

$$p_{n,m}(1) = \frac{m!(n-m)!}{n!} \cdot m \cdot \frac{(n-m)!}{(n-2m+1)!(m-1)!} = \frac{m^2}{n} a_{m,n} \tag{2.9}$$

where

$$a_{m,n} = \frac{(n-m)!}{(n-1)!}\frac{(n-m)!}{(n-2m+1)!} = \frac{(n-m)(n-m-1)\cdots(n-2m+2)}{(n-1)(n-2)\cdots(n-m+1)}.$$

Because the kernel degree m is fixed, we have $a_{m,n} \to 1$ as $n \to \infty$. It follows that

$$\operatorname{var} U_{n,Z}^m \sim \left(m^2/n\right) \zeta_{m,1}, \qquad \text{as } n \to \infty, \tag{2.10}$$

where $x_n \sim y_n$ signifies that the ratio x_n/y_n converges to 1 as $n \to \infty$.

Recall that $\theta = \mathbb{E}[K_m(Z_1, \ldots, Z_m)]$. In the light of (2.10), define the Hájek projection of $U_{n,Z}^m - \theta$ as

$$\begin{aligned}
\hat{U}_{n,Z}^m &:= \sum_{j=1}^n \sum_{I \subset [n], |I|=m} \binom{n}{m}^{-1} \mathbb{E}\left[K_m(Z_I) - \theta \mid Z_j\right] \\
&= \sum_{j=1}^n \binom{n-1}{m-1} \binom{n}{m}^{-1} \hat{K}_m(Z_j) \\
&= \frac{1}{k} \sum_{j=1}^n \hat{K}_m(Z_j)
\end{aligned} \tag{2.11}$$

where $k = n/m$ and

$$\hat{K}_m(x) = \mathbb{E}\left[K_m(Z_1, \ldots, Z_{m-1}, x)\right] - \theta.$$

Because the asymptotic variance of $U_{n,Z}^m$, i.e., $(m^2/n)\zeta_{m,1}$, is exactly equal to the variance of $\hat{U}_{n,Z}^m$, the asymptotic distribution of $U_{n,Z}^m - \theta$ is equal to the one of $\hat{U}_{n,Z}^m$, i.e., as $n \to \infty$,

$$\frac{\hat{U}_{n,Z}^m}{\sqrt{\operatorname{var} \hat{U}_{n,Z}^m}} - \frac{U_{n,Z}^m - \theta}{\sqrt{\operatorname{var} U_{n,Z}^m}} \xrightarrow{p} 0.$$

The Hájek projection $\hat{U}_{n,Z}^m$ is asymptotically normal by the central limit theorem, and thus

$$\frac{k}{\sqrt{n\zeta_{m,1}}} \left(U_{n,Z}^m - \theta\right) \xrightarrow{d} N(0,1), \qquad n \to \infty,$$

provided that the kernel is square-integrable and $\zeta_{m,1} > 0$. For more details, see the proof of Theorem 12.3 in [19]. Note again that we assumed that m is fixed.

Letting $m \to \infty$: first attempt. We aim to adapt the above proof to an extreme U-statistic, that is, when $m \to \infty$ in such a way that $m = \mathrm{o}(n)$ as $n \to \infty$. Such an adaptation is non-trivial because $\zeta_{m,1}$ varies as $m \to \infty$ and the combinatorial numbers $p_{n,m}(\ell)$ no longer satisfy equation (2.8) in general, which makes it difficult to establish the key variance approximation in (2.10).

If we consider again the first term $p_{n,m}(1)$ in (2.9), we see that a further restriction on the range of m is necessary to ensure $a_{m,n} \to 1$ as $n \to \infty$. Clearly,

$$\left(1 - \frac{m-1}{n-m+1}\right)^{m-1} \leq a_{m,n} \leq \left(1 - \frac{m-1}{n-1}\right)^{m-1}. \tag{2.12}$$

The lower and upper bound tend to 1 if and only if $m^2 = \mathrm{o}(n)$ as $n \to \infty$. The condition $m^2 = \mathrm{o}(n)$ is also necessary to ensure that $p_{n,m}(1)$ is the dominant term in the sum $\sum_{\ell=1}^{m} p_{n,m}(\ell)$ as $n \to \infty$. Taken together, these facts lead to the variance approximation (2.10) but only under the restriction $m^2 = \mathrm{o}(n)$ as $n \to \infty$.

Case $m \to \infty$: novel approach. We overcome the restriction $m^2 = \mathrm{o}(n)$ and obtain the desired asymptotic equivalence relation in (2.10) by not truncating Hoeffding's decomposition (2.7) and the following analysis of $\zeta_{m,\ell}$. For integers a and b such that $a \le b$ write $[a{:}b] = [a,b] \cap \mathbb{Z} = \{a, \ldots, b\}$.

Without loss of generality, let the two blocks in (2.7) be $S = [1{:}m]$ and $S' = [(m-\ell+1){:}(2m-\ell)]$. Their overlap $S \cap S' = [(m-\ell+1){:}m]$ contains exactly ℓ elements. It follows that, for $\ell \in [1{:}m]$,

$$\zeta_{m,\ell} = \mathbb{E}\left[K_m(Z_{[1:m]})\, K_m(Z_{[(m-\ell+1):(2m-\ell)]})\right] - \theta^2.$$

The dependence between $K_m(Z_S)$ and $K_m(Z_{S'})$ stems from observations Z_i with $i \in S \cap S'$ in the top-q of both Z_S and $Z_{S'}$. If ℓ is small compared to m, then usually there will be at most one such observation. We can thus expect that $\zeta_{m,\ell}$ can be approximated by $\ell\zeta_{m,1}$. Recognizing the term $p_{n,m}(\ell)$ as the probability mass function of a hypergeometric random variable, we use the expectation formula to find

$$\sum_{\ell=1}^{m} p_{n,m}(\ell)\ell\zeta_{m,1} = \frac{m^2}{n}\zeta_{m,1} = \frac{m\zeta_{m,1}}{k}$$

where $k = n/m$. We obtain the bound

$$\left|\operatorname{var} U_{n,Z}^{m} - \frac{m\zeta_{m,1}}{k}\right| \le \sum_{\ell=1}^{m} p_{n,m}(\ell)\left|\zeta_{m,\ell} - \ell\zeta_{m,1}\right|. \tag{2.13}$$

In order to formally derive (2.10) as well as the asymptotic distribution of the extreme U-statistic $U_{n,Z}^{m}$, we show three limit relations. First, we show that the bound in (2.13) is $\mathrm{o}(1/k)$ as $n \to \infty$, that is,

$$\lim_{n\to\infty} \left|k \operatorname{var} U_{n,Z}^{m} - m\zeta_{m,1}\right| = 0. \tag{2.14}$$

Second, we show that

$$\lim_{m\to\infty} m\zeta_{m,1} = \sigma^2 > 0. \tag{2.15}$$

Note then that (2.14) and (2.15) imply (2.10) as well as

$$\frac{\operatorname{var} U_{n,Z}^{m}}{\operatorname{var} \hat{U}_{n,Z}^{m}} = \frac{\operatorname{var} U_{n,Z}^{m}}{(n/k^2)\zeta_{m,1}} = \frac{k \operatorname{var} U_{n,Z}^{m}}{m\zeta_{m,1}} \to 1, \qquad \text{as } n \to \infty.$$

Finally, we show the asymptotic normality of the Hájek projection $\hat{U}_{n,Z}^{m}$, which, in the extreme value setting, is a centered and reduced row sum of a triangular array of row-wise independent and identically distributed random variables.

Limit relation (2.14) is the content of Proposition 2.3. Limit relation (2.15) and the asymptotic normality of the Hájek projection $\hat{U}^m_{n,Z}$ are the content of Proposition 2.4. Together, Propositions 2.3 and 2.4 lead to the main result of this section in Theorem 2.5.

Conditions and results. We start by listing the required conditions on the kernel K and the degree sequence $m := m_n$. Condition 1 is an integrability condition on the kernel K to be checked by analysis.

Condition 1. *The kernel K is location-scale invariant and not everywhere constant, implying $q \geq 3$. Moreover, for all $p > 0$ it holds that*

$$\mathbb{E}\big[\big|K\big((Z_{q-j:q-1})_{j=1}^{q}\big)\big|^p\big] < \infty,$$

for independent $\mathrm{GP}(\gamma)$ random variables $Z_1, \ldots, Z_{q-1}$ and $Z_{0:q-1} := 0$.

Condition 2. *The degree sequence $m := m_n$ satisfies $m = \mathrm{o}(n^{1-\epsilon})$ and $\ln(n) = \mathrm{o}(m^{1-\delta})$ as $n \to \infty$ for some $\epsilon, \delta > 0$.*

We obtain the following propositions.

Proposition 2.3. *If Conditions 1 and 2 are satisfied, we have*

$$\left|\operatorname{var} U_n^m - \frac{m}{k}\zeta_{m,1}\right| = \mathrm{o}(1/k), \qquad n \to \infty.$$

Let $\overline{K}_m = K_m - \theta$, with θ in (2.5), denote the centred kernel. Recall that for integer $q \geq 1$, the $\mathrm{Erlang}(q)$ distribution is the one of the sum $E_1 + \cdots + E_q$ of q independent unit exponential random variables.

Proposition 2.4. *If Condition 1 holds, then*

$$\lim_{m\to\infty} m\zeta_{m,1} = \int_0^\infty \left(\mathbb{E}\left[\overline{K}_{q+1}\big((Z_{q-j:q-1})_{j=1}^{q}, h_\gamma(S_q/x)\big)\right]\right)^2 \mathrm{d}x := \sigma_K^2(\gamma), \tag{2.16}$$

where S_q is an $\mathrm{Erlang}(q)$ random variable, independent of the independent $\mathrm{GP}(\gamma)$ random variables $Z_1, \ldots, Z_{q-1}$, and where $Z_{0:q-1} := 0$. If, additionally, Condition 2 holds, the Hájek projection $\hat{U}^m_{n,Z}$ in (2.11) satisfies

$$\sqrt{k}\hat{U}^m_{n,Z} \xrightarrow{d} N\left(0, \sigma_K^2(\gamma)\right), \qquad n \to \infty. \tag{2.17}$$

The proofs of Propositions 2.3 and 2.4 are long and are given in Sections A and B, respectively. Together, the two propositions are the required ingredients for deriving the asymptotic distribution of $U^m_{n,Z}$. Since $\hat{U}^m_{n,Z}$ is an orthogonal linear projection of $U^m_{n,Z} - \theta$ and since $\operatorname{var}\hat{U}^m_{n,Z}/\operatorname{var} U^m_{n,Z} \to 1$ as $n \to \infty$, it follows by Theorem 11.2 in [19] that also

$$\frac{U^m_{n,Z} - \theta}{\sqrt{\operatorname{var} U^m_{n,Z}}} - \frac{\hat{U}^m_{n,Z}}{\sqrt{\operatorname{var}\hat{U}^m_{n,Z}}} \xrightarrow{p} 0, \qquad n \to \infty,$$

which proves the main theorem of this section.

Theorem 2.5. *If Conditions 1–2 are satisfied, then the extreme U-statistic $U_{n,Z}^m$ in (2.6) satisfies*

$$\sqrt{k}\left(U_{n,Z}^m - \theta\right) \xrightarrow{d} N\left(0, \sigma_K^2(\gamma)\right), \qquad n \to \infty,$$

for $\theta = \mu(\gamma)$ in (2.5) and $\sigma_{K,\gamma}^2$ in (2.16).

3. Asymptotic theory: sample drawn from the DoA of a GEV distribution

Consider again the U-statistic U_n^m in (1.1) but, in contrast to Section 2, let $X_1, \ldots, X_n$ be a random sample with common distribution function F that satisfies the DoA condition (1.3) with shape parameter $\gamma \in \mathbb{R}$. Our goal now is to prove the asymptotic normality (2.1) in full generality, not only for generalized Pareto distributed samples. The technique relies on a coupling construction, leading to a $\mathrm{GP}(\gamma)$ sample $Z_1, \ldots, Z_n$ associated to $X_1, \ldots, X_n$. The extreme U-statistic $U_{n,Z}^m$ of the $\mathrm{GP}(\gamma)$ sample will be close to U_n^m, up to an asymptotic bias term.

Recall the tail quantile function U of F in (2.2). To establish asymptotic theory, it will be convenient to construct the i.i.d. sample as

$$(X_1, \ldots, X_n) = (U(Y_1^*), \ldots, U(Y_n^*)),$$

for independent standard Pareto random variables $Y_1^*, \ldots, Y_n^*$.

Since the distribution function F is in the DoA of a GEV distribution with shape parameter $\gamma \in \mathbb{R}$, there exists a measurable auxiliary function $a : (1, \infty) \to (0, \infty)$ such that

$$\lim_{x \to \infty} \frac{U(xy) - U(x)}{a(x)} = h_\gamma(y), \qquad \text{locally uniformly in } y \in (0, \infty), \tag{3.1}$$

with h_γ the function in (2.3); see for instance [4, Theorem 1.1.6]. Define

$$(Z_1, \ldots, Z_n) = (h_\gamma(Y_1^*), \ldots, h_\gamma(Y_n^*)).$$

Representation (2.4), relation (3.1), the location-scale invariance of K and a suitable integrability condition then imply

$$\begin{aligned}
&\lim_{n \to \infty} \mathbb{E}[K_m(X_1, \ldots, X_m)] \\
&= \lim_{m \to \infty} \mathbb{E}\left[K\left(\left(\frac{U(Y_{q-j:q-1} Y_{m-q+1:m}^*) - U(Y_{m-q+1:m}^*)}{a(Y_{m-q+1:m}^*)}\right)_{j=1}^q\right)\right] \\
&= \mathbb{E}\left[K\left((h_\gamma(Y_{q-j:q-1}))_{j=1}^q\right)\right] = \mathbb{E}\left[K\left((Z_{q-j:q-1})_{j=1}^q\right)\right],
\end{aligned} \tag{3.2}$$

where $Z_{0:q-1} := 0$. The construction of $(Z_i)_{i=1}^n$ and $(X_i)_{i=1}^n$ on the same probability space allows to find the bias component $B_K(\rho, \gamma)$ in (2.1). The value of the latter depends on how fast the convergence takes place in (3.2), which in turn also depends on how fast the convergence takes place in (3.1).

As in Section 2, denote the U-statistic corresponding to the i.i.d. random variables $Z_1, \dots, Z_n$ with a GP(γ) distribution by

$$U^m_{n,Z} = \binom{n}{m}^{-1} \sum_{I \subset [n], |I|=m} K_m(Z_I).$$

We will prove that as $n \to \infty$,

$$\sqrt{k}\left(U^m_n - U^m_{n,Z}\right) \xrightarrow{p} \lambda B_K(\rho, \gamma), \tag{3.3}$$

where $B_K(\rho,\gamma)$ and $\lambda \geq 0$ will be defined below. Then, the asymptotic distribution of $\sqrt{k}\,(U^m_n - \theta)$ immediately follows from Theorem 2.5. This is the content of Section 3.1, culminating in the paper's main result, Theorem 3.1. In Section 3.2, we consider the special but important case where the kernel degree is $q = 3$, leading to some simplified conditions.

3.1. Asymptotic normality

To prove (3.3), the following second-order conditions are required.

Condition 3. *The tail quantile function $U(t) = \inf\{y : (1 - F(y))^{-1} \geq t\}$ is twice differentiable. Furthermore, the function A defined by*

$$A(t) = \frac{t\,U''(t)}{U'(t)} - \gamma + 1$$

is eventually positive or negative for large t and satisfies $\lim_{t\to\infty} A(t) = 0$, while the function $|A|$ is regularly varying with index $\rho \leq 0$.

Under Condition 3, following Theorem 2.3.12 in [4], we have, for $x > 0$,

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{t\,U'(t)} = h_\gamma(x),$$

and

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{tU'(t)} - h_\gamma(x)}{A(t)} = \int_1^x s^{\gamma-1} \int_1^s u^{\rho-1}\,\mathrm{d}u\,\mathrm{d}s =: H_{\gamma,\rho}(x).$$

Condition 4. *The sequence $k := n/m$ satisfies $\sqrt{k}A(m) \to \lambda \geq 0$ as $n \to \infty$.*

Relation (3.3) can now be proved via the construction of another U-statistic. Recall that I is a set of m positive integers. Define

$$\xi^I_m = \frac{K_m(X_I) - K_m(Z_I)}{A(m)}$$

for A as in Condition 3 and simply write ξ_m whenever only the marginal distribution is of concern. We find

$$\begin{aligned}\sqrt{k}\left(U_n^m - U_{n,Z}^m\right) &= \sqrt{k}\binom{n}{m}^{-1} \sum_{I\subset[n],|I|=m} \left(K_m\left(X_I\right) - K_m\left(Z_I\right)\right) \\ &= \sqrt{k}A(m)\, U_{n,\xi}^m,\end{aligned}$$

where, for fixed m,

$$U_{n,\xi}^m = \binom{n}{m}^{-1} \sum_{I\subset[n],|I|=m} \xi_m^I$$

is another U-statistic. By Hoeffding's decomposition,

$$\operatorname{var}\left(U_{n,\xi}^m\right) = \sum_{\ell=1}^m p_{n,m}(\ell)\Psi_{m,\ell}, \qquad \text{with } p_{n,m}(\ell) = \binom{n}{m}^{-1}\binom{m}{\ell}\binom{n-m}{m-\ell}$$

and $\Psi_{m,\ell} = \operatorname{cov}(\xi_m^I, \xi_m^{I'})$ for index sets $I, I' \subset [n]$ such that $|I'| = |I| = m$ and $|I \cap I'| = \ell$. If it holds that

$$\sum_{\ell=1}^m p_{n,m}(\ell)\Psi_{m,\ell} \to 0, \qquad n\to\infty, \tag{3.4}$$

then we have by Chebychev's inequality that

$$U_{n,\xi}^m - \mathbb{E}[\xi_m] \xrightarrow{p} 0, \qquad n\to\infty.$$

If moreover

$$\lim_{m\to\infty} \mathbb{E}[\xi_m] = \mathbb{E}\left[\lim_{m\to\infty} \xi_m\right] =: B_K(\rho,\gamma), \tag{3.5}$$

then clearly also

$$U_{n,\xi}^m \xrightarrow{p} B_K(\rho,\gamma), \qquad n\to\infty.$$

From the decomposition

$$\sqrt{k}(U_n^m - \theta) = \sqrt{k}\left(U_n^m - U_{n,Z}^m\right) + \sqrt{k}\left(U_{n,Z}^m - \theta\right),$$

together with Theorem 2.5, we obtain the following theorem.

Theorem 3.1. *Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with a common distribution function F that satisfies the DoA condition (1.3) with shape parameter $\gamma \in \mathbb{R}$. Under Conditions 1-4 and provided relations (3.4) and (3.5) are met, the corresponding extreme U-statistic, U_n^m, satisfies*

$$\sqrt{k}\left(U_n^m - \theta\right) \xrightarrow{d} N(\lambda B_K(\rho,\gamma), \sigma_K^2(\gamma)),$$

where $\sigma_K^2(\gamma)$ is as in Proposition 2.4 and $\theta = \mu(\gamma)$ is as in (2.5) .

Note that $\sigma_K^2(\gamma)$ does not depend on the distribution of X. Therefore, a practical consequence of Theorem 3.1 is that, for kernels K satisfying the required conditions, one can do a parametric bootstrap using $GP(\hat{\gamma})$ distributed random variables to obtain an estimate of $\sigma_K^2(\gamma)$, where $\hat{\gamma}$ can be any reasonable estimator of γ, provided that $\sigma_K^2(\gamma)$ is a continuous function of γ.

To make use of Theorem 3.1, one needs to verify that the kernel K satisfies relations (3.4) and (3.5). In Section 3.2, we deal with kernels where $q = 3$ — the minimum q required to obtain a location-scale invariant kernel that is not everywhere constant— and proceed to give conditions under which relation (3.5) is satisfied. Relation (3.4) will follow as a by-product.

3.2. Location-scale invariant kernels where $q = 3$

For a location-scale invariant kernel $K : \mathbb{R}^3 \to \mathbb{R}$, we may write

$$\begin{aligned} K(x_1, x_2, x_3) = K(x_1 - x_2, 0, x_3 - x_2) &= K\left(\frac{x_1 - x_2}{x_2 - x_3}, 0, -1\right) \\ &=: g\left(\ln\left(\frac{x_1 - x_2}{x_2 - x_3}\right)\right) \end{aligned}$$

for some function $g : \mathbb{R} \to \mathbb{R}$, with $x_1 > x_2 > x_3$. Condition 3 implies that the distribution function F is continuous, so that ties do not occur with probability one and we do not need to worry about the cases $x_1 = x_2$ or $x_2 = x_3$. If the kernel K is differentiable and the first-order partial derivatives $\dot{K}_j(x) = \partial K(x)/\partial x_j$ for $j \in [3]$ are continuous, then g must be differentiable and we have

$$\begin{aligned} \dot{K}_1(x) &= \frac{g'\left(\ln\left(\frac{x_1 - x_2}{x_2 - x_3}\right)\right)}{x_1 - x_2}, \\ \dot{K}_2(x) &= -\frac{g'\left(\ln\left(\frac{x_1 - x_2}{x_2 - x_3}\right)\right)}{x_1 - x_2} - \frac{g'\left(\ln\left(\frac{x_1 - x_2}{x_2 - x_3}\right)\right)}{x_2 - x_3}, \\ \dot{K}_3(x) &= \frac{g'\left(\ln\left(\frac{x_1 - x_2}{x_2 - x_3}\right)\right)}{x_2 - x_3}, \end{aligned}$$

where g' is the derivative of g. We assume the following with respect to g'.

Condition 5. *The kernel K is differentiable. Moreover, g' is monotone and there exists $\epsilon > 0$ such that*

$$\limsup_{m \geq 3} \mathbb{E}\left[\left|g'\left(\ln\left(\frac{X_{m:m} - X_{m-1:m}}{X_{m-1:m} - X_{m-2:m}}\right)\right)\right|^{2+\epsilon}\right] < \infty,$$

or, alternatively, g' is bounded (but not necessarily monotone).

We remark that kernels that are a linear combination of log order statistic spacings have bounded auxiliary function g', see Section 4.

In addition, we assume that the moments of the kernel function do not grow too quickly as $n \to \infty$.

Condition 6. *There exists some $\epsilon > 0$ and $a > 0$ such that*

$$\mathbb{E}[|K_m(X_1, \dots, X_m)|^{2+\epsilon}] = \mathrm{O}(m^a), \qquad m \to \infty. \tag{3.6}$$

We obtain the following corollary of Theorem 3.1.

Corollary 3.2. *For a kernel $K : \mathbb{R}^3 \to \mathbb{R}$, under Conditions 3–6, relations (3.4) and (3.5) are satisfied, which implies*

$$\sqrt{k}\left(U_n^m - U_{n,Z}^m\right) \xrightarrow{p} \lambda B_K(\rho, \gamma), \qquad n \to \infty,$$

where

$$B_K(\rho, \gamma) = \mathbb{E}\left[S_3^{-\rho}\right] \mathbb{E}\left[\sum_{j=1}^{2} \dot{K}_j\left(\{h_\gamma\left(Y_{3-j:2}\right)\}_{j=1}^3\right) H_{\gamma,\rho}(Y_{3-j:2})\right], \tag{3.7}$$

and S_3 is an Erlang(3) *random variable. If, additionally, Conditions 1 and 2 hold, then by Theorem 3.1, the extreme U-statistic U_n^m satisfies*

$$\sqrt{k}\left(U_n^m - \theta\right) \xrightarrow{d} N\left(\lambda B_K(\rho, \gamma), \sigma_K^2(\gamma)\right), \qquad n \to \infty$$

where $\sigma_K^2(\gamma)$ is as in Proposition 2.4 and $\theta = \mu(\gamma)$ is as in (2.5).

A similar theory could be developed for a scale invariant kernel K with $q = 2$. For example, for $\gamma > 0$ and such a kernel, one needs to impose integrability conditions to ensure

$$\lim_{m\to\infty} \mathbb{E}\left[K_m(X_1, \dots, X_m)\right] = \lim_{m\to\infty} \mathbb{E}\left[K(X_{m:m}/X_{m-1:m}, 1)\right] = \mathbb{E}\left[K(Y_1^\gamma, 1)\right].$$

For an illustration of the case $q = 1$ where neither scale nor location invariance of the kernel function can be exploited, see [14].

4. Estimating the extreme value index

In this section, we develop a kernel that results in an unbiased estimator of the shape parameter $\gamma \in \mathbb{R}$ when the observations are independently sampled from a GP(γ) distribution. The new kernel is built out of the location-scale invariant kernels

$$\begin{aligned} K_{m,1}(x_1, \dots, x_m) &= \ln\left(\frac{x_{m:m} - x_{m-1:m}}{x_{m-1:m} - x_{m-2:m}}\right), \\ K_{m,2}(x_1, \dots, x_m) &= \ln\left(\frac{x_{m:m} - x_{m-2:m}}{x_{m-1:m} - x_{m-2:m}}\right). \end{aligned} \tag{4.1}$$

Kernel $K_{m,1}$ keeps the so called Galton ratios as small as possible, see [17], while $K_{m,2}$ was considered in [16]. Kernels such as $K_{m,1}$ and $K_{m,2}$ are, in some sense, in canonical form because in Section 3.2 we have seen that any location-scale invariant kernel $K : \mathbb{R}^q \to \mathbb{R}$, for $q = 3$, has representation

$$g\left(\ln\left(\frac{x_1 - x_2}{x_2 - x_3}\right)\right), \qquad \text{for } x_1 > x_2 > x_3.$$

For an independent random sample $X_1, \ldots, X_m$ from a distribution F with extreme value index γ, we have, under suitable integrability conditions on F,

$$\lim_{m\to\infty} \mathbb{E}\left[\ln\left(\frac{X_{m:m} - X_{m-1:m}}{X_{m-1:m} - X_{m-2:m}}\right)\right] = \mathbb{E}\left[\ln\left(\frac{Z_{2:2} - Z_{1:2}}{Z_{1:2}}\right)\right] =: u_1(\gamma),$$

and

$$\lim_{m\to\infty} \mathbb{E}\left[\ln\left(\frac{X_{m:m} - X_{m-2:m}}{X_{m-1:m} - X_{m-2:m}}\right)\right] = \mathbb{E}\left[\ln\left(\frac{Z_{2:2}}{Z_{1:2}}\right)\right] =: u_2(\gamma),$$

where $Z_{2:2} \geq Z_{1:2}$ are $\mathrm{GP}(\gamma)$ order statistics. In what follows we construct a fixed linear combination of $u_1(\gamma)$ and $u_2(\gamma)$ that is equal to γ.

To calculate expectations such as $u_1(\gamma)$ and $u_2(\gamma)$ we use known formulas for the expectation of exponentiated Generalized Pareto random variables and their order statistics in terms of the digamma function ψ. The proof of the following lemma is given in Section D.

Lemma 4.1. *Let $\gamma \in \mathbb{R}$ and let Z_1, Z_2 be two independent $\mathrm{GP}(\gamma)$ random variables, with order statistics $Z_{1:2} \leq Z_{2:2}$. We have*

$$\mathbb{E}[\ln Z_1] = \begin{cases} \ln(1/\gamma) + \psi(1) - \psi(1/\gamma) & \text{if } \gamma > 0 \\ \ln(-1/\gamma) + \psi(1) - \psi(1 - 1/\gamma) & \text{if } \gamma < 0 \\ \psi(1) & \text{if } \gamma = 0, \end{cases}$$

$$\mathbb{E}[\ln Z_{2:2}] = \begin{cases} \ln(1/\gamma) + \psi(1) + \psi(2/\gamma) - 2\psi(1/\gamma) & \text{if } \gamma > 0 \\ \ln(-1/\gamma) + \psi(1) + \psi(1 - 2/\gamma) - 2\psi(1 - 1/\gamma) & \text{if } \gamma < 0 \\ \psi(1) + \ln(2) & \text{if } \gamma = 0, \end{cases}$$

$$\mathbb{E}[\ln Z_{1:2}] = \begin{cases} \ln(1/\gamma) + \psi(1) - \psi(2/\gamma) & \text{if } \gamma > 0 \\ \ln(-1/\gamma) + \psi(1) - \psi(1 - 2/\gamma) & \text{if } \gamma < 0 \\ \psi(1) - \ln(2) & \text{if } \gamma = 0. \end{cases}$$

As a consequence, for all $\gamma \in \mathbb{R}$, we have

$$2u_1(\gamma) - u_2(\gamma) = \gamma. \tag{4.2}$$

Lemma 4.1 motivates the definition of the kernel function

$$K_m^* = 2K_{m,1} - K_{m,2}$$

with $K_{m,1}$ and $K_{m,2}$ as in (4.1), or more specifically,

$$K_m^*(x_1, \ldots, x_m) = K^*(x_{m:m}, x_{m-1:m}, x_{m-2:m}), \qquad (x_1, \ldots, x_m) \in \mathbb{R}^m,$$

with K^* the permutation invariant function on $\mathbb{R}^3$ given by

$$K^*(y_1, y_2, y_3) = 2\ln\left(\frac{y_1 - y_2}{y_2 - y_3}\right) - \ln\left(\frac{y_1 - y_3}{y_2 - y_3}\right), \qquad y_1 > y_2 > y_3. \tag{4.3}$$

Equation (4.2) implies that the extreme U-statistic based on K^* has a particularly attractive property.

Corollary 4.2. *For $\gamma \in \mathbb{R}$, integer $n \geq m \geq 3$, and independent* $\mathrm{GP}(\gamma)$ *random variables $Z_1, \ldots, Z_n$, we have*

$$\mathbb{E}\left[K_m^*(Z_1, \ldots, Z_m)\right] = \gamma.$$

Consequently, the U-statistic

$$U_{n,*,Z}^m := \binom{n}{m}^{-1} \sum_{I \subset [n], |I| = m} K_m^*(Z_I)$$

is an unbiased estimator of γ.

For a random sample $X_1, \ldots, X_m$ from a distribution F with extreme value index γ, the kernel $K_m^*(X_1, \ldots, X_m)$ is not necessarily an unbiased estimator of γ, but may still satisfy

$$\lim_{m \to \infty} \mathbb{E}\left[K_m^*(X_1, \ldots, X_m)\right] = 2u_1(\gamma) - u_2(\gamma) = \gamma,$$

provided a suitable integrability condition holds. Corollary 3.2 then implies the asymptotic normality of $\sqrt{k}(U_{n,*}^m - \gamma)$ where

$$U_{n,*}^m = \binom{n}{m}^{-1} \sum_{I \subset [n], |I| = m} K_m^*(X_I). \tag{4.4}$$

A weak sufficient condition on F is stated next.

Proposition 4.3. *The kernel K_m^* satisfies all required integrability conditions in Corollary 3.2, except possibly Condition 6.*

Let the distribution function F have extreme value index $\gamma \in \mathbb{R}$. If F admits a density function f with support $\mathcal{E}$ and if there exists $\epsilon > 0$ and $p > 2$ such that both $\int_{(-\infty,-1)} (\ln(-x))^p \,\mathrm{d}F(x) < \infty$ and

$$\sup_{x \in \mathcal{E}} \left| F(x) - F(x + e^{-r^{1/p}}) \right| = \mathrm{O}(r^{-(1+\epsilon)}), \qquad \text{as } r \to \infty, \tag{4.5}$$

then Condition 6 is satisfied. In particular, (4.5) holds if the density f is bounded.

By Corollary 3.2, we conclude that under the sufficient conditions in Proposition 4.3, the extreme U-statistic $U_{n,*}^m$ in (4.4) is an asymptotically normal estimator for γ:

$$\sqrt{k}(U_{n,*}^m - \gamma) \xrightarrow{d} N\left(\lambda B_{K^*}(\gamma, \rho), \sigma_{K^*}^2(\gamma)\right), \qquad n \to \infty,$$

The asymptotic variance $\sigma^2_{K^*}$ and bias B_{K^*} are given in equations (2.16) and (3.7), respectively, with K to be replaced by K^* in (4.3).

The finite-sample performance of $U^m_{n,*}$ as an estimator of γ will be explored in Section 5 by a simulation study. For calculation purposes, it is convenient to express $U^m_{n,*}$ explicitly as a linear combination of the log-spacings.

Lemma 4.4. *We can express the U-statistic $U^m_{n,*}$ based on a random sample $X_1, \ldots, X_n$ in terms of its order statistics as*

$$U^m_{n,*} = \sum_{j=2}^{n-m+3} \frac{\binom{n-j}{m-3}}{\binom{n}{m}} \left(2\frac{n-j-m+3}{m-2} - j \right) \sum_{i=1}^{j-1} \ln(X_{n-i+1:n} - X_{n-j+1:n}). \tag{4.6}$$

For given n and m, the coefficients on the right-hand side of (4.6) lend themselves well for recursive calculation, without the need of invoking factorials:

$$\frac{\binom{n-2}{m-3}}{\binom{n}{m}} = \frac{m(m-1)(m-2)}{n(n-1)(n-m+1)},$$
$$\frac{\binom{n-j}{m-3}}{\binom{n-(j-1)}{m-3}} = \frac{n-j-m+2}{n-j+1}, \qquad j \in [3{:}(n-m+3)].$$

5. Finite-sample performance

Here we compare the performance of the location-scale invariant unbiased estimator $U^m_{n,*}$ in (4.4) of $\gamma \in \mathbb{R}$ to the location-scale invariant GP Maximum Likelihood (ML) estimator [3]. While the latter is also considered an estimator of $\gamma \in \mathbb{R}$, the algorithm to compute the ML estimator is theoretically constrained to $\gamma > -1$ [8], see also [20].

We compare the performance of both estimators on the basis of mean-squared error (MSE). For that purpose, we simulate 100 datasets of $n = 10\,000$ observations for a given distribution and for each dataset use both estimators to estimate γ over a range of block sizes m. To facilitate comparison between the two estimators we fix the number of tail observations to be used in GP ML estimation to $k = 3n/m$, where m is the block size of the kernel K^*_m. Theoretically, for estimators using the top-3 observations of a disjoint partitioning of the dataset in blocks of size m, the number of data points used is equal to $3n/m$ too; moreover, when $m = 3$, both $U^m_{n,*}$ and the GP ML estimator use all data.

5.1. Random variables with a GP distribution

We start by comparing the MSE of the GP ML estimator to $U^m_{n,*}$ in the case that the data are sampled from the GP distribution with shape parameter $\gamma \in \{-0.5, 0, 0.5\}$. Effectively, the comparison will be on the basis of estimator variance, since, in the GP case, $U^m_{n,*}$ is unbiased and the GP ML estimator is asymptotically unbiased.

Fig 1: MSE comparison of $U_{n,*}^m$ and the GP ML estimator for 100 samples of size $n = 10\,000$ from $\mathrm{GP}(\gamma)$

(a) $\mathrm{GP}(\gamma = 0.5)$ (b) $\mathrm{GP}(\gamma = 0)$ (c) $\mathrm{GP}(\gamma = -0.5)$

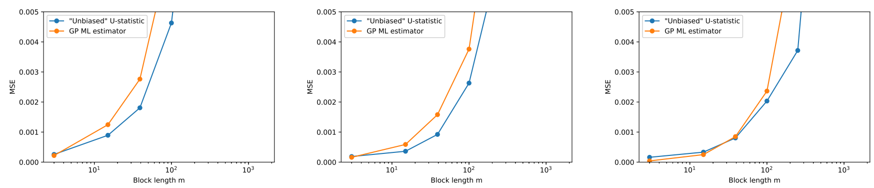


Table 1

Minimal root-mean-squared-error (RMSE) for $n = 10\,000$ $\mathrm{GP}(\gamma)$ *observations*

	$\gamma = -0.5$	$\gamma = 0$	$\gamma = -0.5$
RMSE GP ML	0.0148	0.0401	0.0064
RMSE $U_{n,*}^m$	0.0160	0.0432	0.0127

Figure 1 shows that the MSE for both estimators is minimal when we use all available observations in estimation (block length $m = 3$). Table 1 shows that the minimal MSE of $U_{n,*}^m$ is close to that of the GP ML estimator.

Moreover, the MSE of $U_{n,*}^m$ is significantly lower than that of the GP ML estimator over the largest range of the block length m, except when $\gamma = -0.5$. The apparent discrepancy between the two cases $\gamma \in \{0, 0.5\}$ on the one hand and the case $\gamma = -0.5$ on the other hand can be understood by a comparison of the asymptotic variance.

Recall that the GP ML estimator has asymptotic variance $(1+\gamma)^2$, provided $\gamma > -1/2$ [6]. For the asymptotic variance of $U_{n,*}^m$, we fix a grid of 50 equally spaced values of γ between -1 and 1 and approximate $\sigma_{K^*}^2(\gamma)$ by the sample variance of $U_{n,*}^m$ based on 1000 $\mathrm{GP}(\gamma)$ samples of size $n = 10\,000$, multiplied with $k = n/m = 100$. The normalized asymptotic variance of the GP ML estimator $(1+\gamma)^2/3$ (like before, we divide by three to facilitate comparison) together with the estimated asymptotic variance $\hat{\sigma}_{K^*}^2(\gamma)$ corresponding to $U_{n,*}^m$ are depicted in Figure 2. The asymptotic variances intersect around $\gamma \approx -0.25$, with $\sigma_{K^*}^2(\gamma)$ indeed being higher for lower values of γ but lower for all higher values of γ. In this context it is possible for the asymptotic variance $\sigma_{K^*}^2(\gamma)$ to be lower than the one of the GP ML estimator because $U_{n,*}^m$ uses more observations than the GP ML estimator. For completeness the approximate values of $\sigma_{K^*}^2(\gamma)$ for $\gamma \in [-1, 1]$ can be found in Table 2.

Fig 2: Normalized asymptotic variance $\sigma^2_{K^*}(\gamma)$ of $U^m_{n,*}$ and $(1+\gamma)^2/3$, corresponding to the one of the GP ML estimator

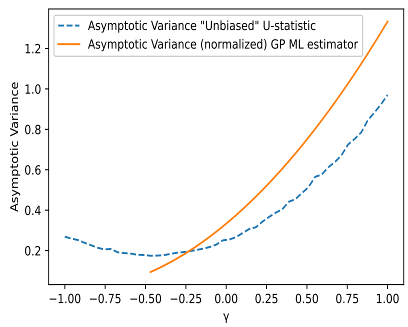
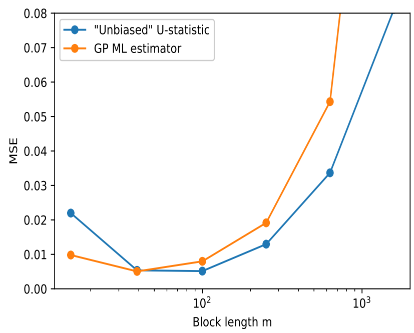


Table 2
Asymptotic variance $\sigma^2_{K^*}(\gamma)$ for $\gamma \in [-1, 1]$

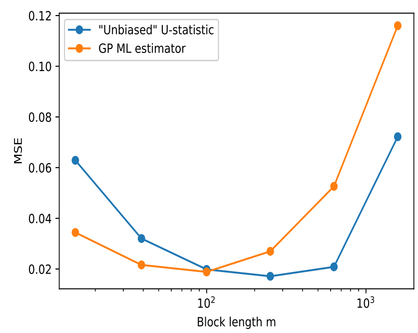
γ	$\sigma^2_{K^*}(\gamma)$	γ	$\sigma^2_{K^*}(\gamma)$	γ	$\sigma^2_{K^*}(\gamma)$	γ	$\sigma^2_{K^*}(\gamma)$	γ	$\sigma^2_{K^*}(\gamma)$
−1.000	0.269	−0.959	0.259	−0.918	0.252	−0.878	0.239	−0.837	0.227
−0.796	0.214	−0.755	0.206	−0.714	0.208	−0.673	0.192	−0.633	0.188
−0.592	0.185	−0.551	0.179	−0.510	0.178	−0.469	0.175	−0.429	0.175
−0.388	0.177	−0.347	0.183	−0.306	0.189	−0.265	0.193	−0.224	0.198
−0.184	0.203	−0.143	0.210	−0.102	0.218	−0.061	0.230	−0.020	0.251
0.020	0.255	0.061	0.267	0.102	0.287	0.143	0.307	0.184	0.316
0.225	0.343	0.265	0.366	0.306	0.387	0.347	0.404	0.388	0.442
0.429	0.454	0.469	0.485	0.510	0.514	0.551	0.565	0.592	0.576
0.633	0.612	0.674	0.638	0.714	0.672	0.755	0.725	0.796	0.752
0.837	0.785	0.878	0.845	0.918	0.883	0.959	0.924	1.000	0.970

Fig 3: MSE comparison of $U_{n,*}^m$ and the GP ML estimator for 100 samples of size $n = 10\,000$ from the stated distributions, with extreme value index $\gamma \in \mathbb{R}$ and second-order parameter $\rho \leq 0$ (Condition 3) as indicated

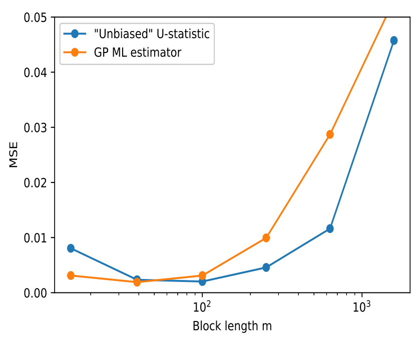
(a) Student-t(2), $\gamma = 0.5$, $\rho = -1$



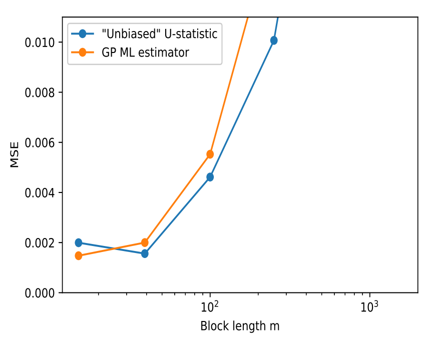
(b) Normal, $\gamma = 0$, $\rho = 0$



(c) Beta(2, 2), $\gamma = -0.5$



(d) Frechét, $\gamma = 1/3$, $\rho = -1$



5.2. *Random variables in the DoA of a GEV distribution*

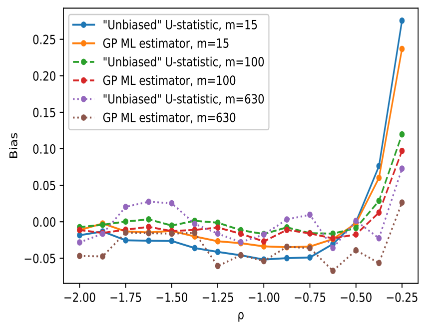
Next we consider data drawn from a distribution function in the DoA of a GEV distribution. The optimal block length m in a MSE sense will then not be equal to three, but will in general be a higher value that better balances bias and variance.

In Figure 3 we illustrate the relative performance of $U_{n,*}^m$ in terms of MSE for a couple of well known distributions. The conclusions from the previous section remain; the minimal MSE of $U_{n,*}^m$ is competitive with that of the GP ML estimator (sometimes even lower) and the performance of $U_{n,*}^m$ is usually better over the largest range of the block length.

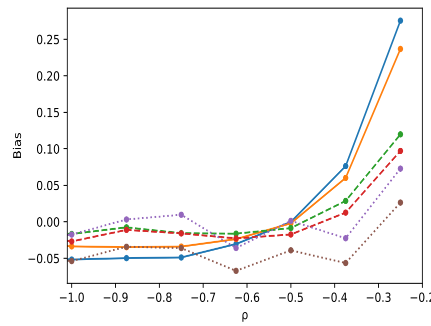
MSE comparisons for observations from a distribution in the DoA of a GEV distribution effectively net out the competing effects of changes in estimator

Fig 4: Bias of $U_{n,*}^m$ and the GP ML estimator based on 1000 samples of size $n = 10\,000$ from the Burr distribution with stated extreme value index γ and second-order parameter ρ

(a) $\gamma = 0.5, \rho \in [-2, -0.25]$



(b) $\gamma = 0.5, \rho \in [-1, -0.25]$



variance (depending on γ) to changes in estimator bias (depending on the second-order parameter ρ as well as on γ). For many distributions, γ and ρ cannot be controlled separately. This is why we study the Burr distribution to disentangle the effects of ρ and γ. The Burr(λ, η) distribution has distribution function $F(x) = 1 - (1/(1 + x^\eta))^\lambda$ for $\eta, \lambda, x > 0$, with extreme value index $\gamma = (\lambda\eta)^{-1}$ and second-order parameter $\rho = -\lambda^{-1}$.

Figure 4 depicts the bias for several choices of m based on 1000 simulated datasets of $n = 10\,000$ observations from a Burr distribution where the parameters are chosen such that $\gamma = 1/2$ is fixed as we vary ρ. As expected, the bias of both estimators approaches 0 as ρ decreases and increases steeply as ρ approaches 0. While subplot (a) of Figure 4 shows that the bias of $U_{n,*}^m$ is somewhat lower than that of the GP ML estimator when ρ is very negative, subplot (b) shows that the bias of $U_{n,*}^m$ is higher uniformly over m when ρ approaches 0.

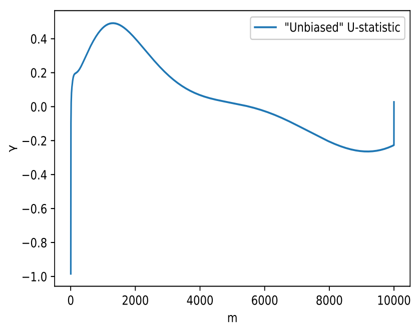
Finally, Figure 5 shows that a single trajectory of the estimator $U_{n,*}^m$ as a function of unit increments in m is smooth, in contrast to the characteristically erratic "Hill plot" of the GP ML estimator as a function of the top k order statistics. Here, no normalizing relation between m and k such as $k = 3n/m$ is used, ensuring that for both estimators one-unit increments can be shown. Nonetheless, choosing the block size in practice remains a difficult task even when the single trajectory of the estimator $U_{n,*}^m$ is smooth.

6. Discussion

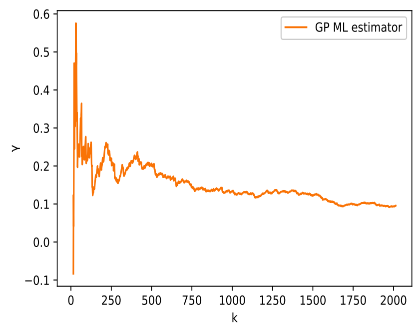
In Theorem 2.5, as well as elsewhere in the paper, we have assumed that the data are independently and identically distributed. Since γ is a parameter belonging to the marginal distribution of the observations, we expect that

Fig 5: Single trajectories of the estimators as a function of m or k, based on a Student-t(4) sample of size $n = 10\,000$

(a) $U^m_{n,*}$, $m \in [3 : 10\,000]$



(b) GP ML estimator, $k \in [15 : 2000]$



Theorem 2.5 can be adapted to the case where the Z_i all follow the same GP distribution but are weakly dependent. If the dependence is weak enough and the sample size is large, one may still find that for a randomly chosen block Z_I the observations are approximately independent such that $\theta \approx \mathbb{E}[K_m(Z_I)]$ and that the approximation $\zeta_{m,\ell} \sim \ell\zeta_{m,1}$ underlying Proposition 2.3 remains valid. The asymptotic variance will be different, however.

Furthermore, in the simulation study, when data were sampled from the GP(γ) distribution, the estimator $U^3_{n,*}$ turned out to have the lowest variance among all unbiased estimators $U^m_{n,*}$ of γ, for $m \geq 3$. Since the estimator $U^3_{n,*}$ is also a function of the entire sample, it is an open question whether the estimator $U^3_{n,*}$ is the uniformly minimum variance unbiased estimator (UMVUE) for γ of the GP(γ) distribution. Because the family of GP(γ) with $\gamma \in \mathbb{R}$ is not in the exponential class this question is left to further research.

Finally, note that a U-statistic may also be calculated based on all exceedances over a high threshold. If the kernel is location-scale invariant, this amounts to applying a U-statistic only to a certain fraction of the largest observations in a sample. Doing so while keeping $m = q$ fixed would yield another kind of extreme U-statistic, perhaps worth studying as well.

Appendix A: Proof of Proposition 2.3

Recall the bound in Section 2.2, given by

$$\left| \operatorname{var} U^m_{n,Z} - \frac{m\zeta_{m,1}}{k} \right| \leq \sum_{\ell=1}^{m} p_{n,m}(\ell) \left| \zeta_{m,\ell} - \ell\zeta_{m,1} \right| . \tag{A.1}$$

Recall that

$$p_{n,m}(\ell) = \binom{n}{m}^{-1} \binom{m}{\ell} \binom{n-m}{m-\ell} \tag{A.2}$$

for integer $1 \le m \le n$ and $0 \le \ell \le m$, which is the probability mass function of the hypergeometric distribution counting the number of successes after m draws without replacement out of an urn containing n balls of which m have the desired colour. To prove Proposition 2.3, we will decompose the sum over $\ell \in [1{:}m]$ in (A.1) into two pieces, $\ell \le \ell_n$ and $\ell > \ell_n$, with $\ell_n = \mathrm{o}(m)$ determined below.

- For $\ell \le \ell_n$, we will show that $|\zeta_{m,\ell} - \ell\zeta_{m,1}|$ is sufficiently small.
- For $\ell > \ell_n$, we use tail bounds on the hypergeometric distribution $p_{n,m}(\ell)$.

We make these statements precise in the proof below by listing a number of properties that, in combination, imply that the bound (A.1) is $\mathrm{o}(1/k)$ as $n \to \infty$.

Proof of Proposition 2.3. By (A.1), we have, for any $b > 0$,

$$\begin{aligned} \left| \operatorname{var} U_n^m - \frac{m}{k}\zeta_{m,1} \right| &\le \max_{\ell \in [1:\ell_n]} \left| \ell^{-1}\zeta_{m,\ell} - \zeta_{m,1} \right| \cdot \sum_{\ell=1}^{\ell_n} p_{n,m}(\ell)\ell \\ &\quad + \max_{\ell \in [1:m]} \ell^{-b} \left| \zeta_{m,\ell} - \ell\zeta_{m,1} \right| \cdot \sum_{\ell=\ell_n+1}^{m} p_{n,m}(\ell)\ell^b. \end{aligned}$$

The expectation of the hypergeometric distribution in (A.2) is $m \cdot m/n = m/k$. The previous decomposition implies the following break-down of (2.14) into three sub-goals: As $n \to \infty$,

$$\max_{\ell \in [1:\ell_n]} \left| \ell^{-1}\zeta_{m,\ell} - \zeta_{m,1} \right| = \mathrm{o}(1/m), \tag{A.3}$$

$$\max_{\ell \in [1:m]} \ell^{-b} \left| \zeta_{m,\ell} - \ell\zeta_{m,1} \right| = \mathrm{O}(1), \tag{A.4}$$

$$\sum_{\ell=\ell_n+1}^{m} p_{n,m}(\ell)\ell^b = \mathrm{o}(1/k), \tag{A.5}$$

where we choose b such that $1/2 \le b \le 1$.

If the above sub-goals are attained, we obtain that

$$\left| \operatorname{var} U_n^m - \frac{m}{k}\zeta_{m,1} \right| \le \mathrm{o}(1/m)m/k + \mathrm{O}(1)\,\mathrm{o}(1/k) = \mathrm{o}(1/k), \qquad \text{as } n \to \infty.$$

The proof of the three sub-goals (A.3), (A.4) and (A.5) for the choice

$$\ell_n = \left\lfloor \left(\frac{m}{n} + \left(\frac{\ln n}{2m} \right)^{1/2} \right) m \right\rfloor, \tag{A.6}$$

with $\lfloor x \rfloor$ the integer part of $x \in \mathbb{R}$, is the content of Lemmas A.1, A.2 and A.3 below. ∎

We remark that the choice of ℓ_n is constrained by (A.3) and (A.5). On the one hand, for (A.3) to hold, we will need that $\ell_n = \mathrm{o}(m)$ as $n \to \infty$. On the other hand, for (A.5) to hold, we will need that $\ell_n \to \infty$ sufficiently fast. Both criteria can be met for the choice of ℓ_n in (A.6), provided m is not too small or large as per Condition 2.

The proof of sub-goal (A.3) is somewhat involved and is developed in Section A.1 via a series of lemmas. First, we establish sub-goals (A.4) and (A.5).

Recall that for the independent $\mathrm{GP}(\gamma)$ random variables $Z_1, \ldots, Z_{2m-1}$,

$$\zeta_{m,\ell} = \mathbb{E}\left[K_m(Z_{[1:m]})\, K_m(Z_{[(m-\ell+1):(2m-\ell)]})\right] - \theta^2,$$

for $\ell \in [1{:}m]$. For integer $m \geq q$ and real $a > 0$, define

$$\|K\|_a := \left(\mathbb{E}\left[\left|K\left((Z_{q-j:q-1})_{j=1}^q\right)\right|^a\right]\right)^{1/a},$$

which, by Lemma 2.2, does not depend on m.

Lemma A.1 (Sub-goal (A.4))**.** *If Condition 1 is satisfied, then, for every integer $m \geq q$, we have, for any $a > 2$,*

$$\zeta_{m,1} \leq (2q)^{1-2/a}\, \|K\|_a^2 \cdot m^{2/a-1}. \tag{A.7}$$

As a consequence, we can take any $b \in (2/a, 1)$ to ensure that (A.4) holds.

Proof. To ensure (A.4), it is sufficient to have

$$\sup_{m \geq q} \max_{\ell \in [1:m]} \zeta_{m,\ell} < \infty, \tag{A.8}$$

$$\sup_{m \geq q} m^{1-b} \zeta_{m,1} < \infty, \tag{A.9}$$

provided $0 \leq b \leq 1$. By Hölder's inequality, Lemma 2.2 and Condition 1,

$$\sup_{m \geq q} \max_{\ell \in [1:m]} \zeta_{m,\ell} \leq \|K\|_2^2 - \theta^2 < \infty,$$

such that (A.8) is guaranteed.

For (A.9), consider the difference

$$\Delta_m = K_m(Z_{[1:m]})\, K_m(Z_{[m:(2m-1)]}) - K_{m-1}(Z_{[1:(m-1)]})\, K_{m-1}(Z_{[(m+1):(2m-1)]}),$$

capturing the effect of omitting Z_m from $Z_1, \ldots, Z_{2m-1}$. It follows that

$$\zeta_{m,1} = \mathbb{E}\left[K_m(Z_{[1:m]}) K_m(Z_{[m:(2m-1)]})\right] - \theta^2 = \mathbb{E}(\Delta_m).$$

Further, if Z_m does not belong to the top-q of $Z_{[1:m]}$ nor to the one of $Z_{[m:(2m-1)]}$, $K_m(Z_{[1:m]}) = K_{m-1}(Z_{[1:(m-1)]})$ and $K_m(Z_{[m:(2m-1)]}) = K_{m-1}(Z_{[(m+1):(2m-1)]})$, which implies that $\Delta_m = 0$. The probability that Z_m belongs to the top-q of

$Z_{[1:m]}$ or $Z_{[m:(2m-1)]}$ is not larger than $2q/m$. We find, by applying Hölder's inequality twice, for any $a > 2$,

$$\begin{aligned}|\mathbb{E}(\Delta_m)| \le \mathbb{E}(|\Delta_m|) &\le \left(\mathbb{E}\left[\left|K_m(Z_{[1:m]})K_m(Z_{[m:(2m-1)]})\right|^{a/2}\right]\right)^{2/a}(2q/m)^{1-2/a} \\ &\le \|K\|_a^2 \cdot (2q/m)^{1-2/a},\end{aligned}$$

which is the upper bound in (A.7). By multiplying the upper bound with m^{1-b} for some $b \in (2/a, 1)$, we obtain (A.4). □

Lemma A.2 (Sub-goal (A.5))**.** *If Condition 2 holds, the sequence ℓ_n in (A.6) satisfies $\ell_n \to \infty$ and $\ell_n = \mathrm{o}(m)$ as $n \to \infty$, and sub-goal (A.5) holds.*

Proof. The statements regarding the asymptotic order of ℓ_n follow immediately. Let H_n be a hypergeometric random variable with probability mass function $p_{n,m}$ in (A.2). We have $\mathbb{E}(H_n) = m/k$. Put $t = (\ell_n + 1)/m - 1/k$ for $t \ge 0$ to be determined. We have

$$\begin{aligned}k\sum_{\ell=\ell_n+1}^{m} p_{n,m}(\ell)\ell^b &\le km^b \sum_{\ell=\ell_n+1}^{m} p_{n,m}(\ell) \\ &= km^b\,\mathbb{P}[H_n \ge (k^{-1}+t)m] \\ &\le \exp\left(-2mt^2 - \ln k^{-1} + b\ln m\right),\end{aligned}$$

see [11] or [2]. Since $t \ge ((2m)^{-1}\ln n)^{1/2}$ and $b < 1$, we get

$$2mt^2 + \ln 1/k - b\ln(m) \ge \ln n - \ln k - b\ln m = (1-b)\ln m \to \infty,$$

as $n \to \infty$. The lemma follows. □

A.1. Proof of sub-goal (A.3)

It remains to prove the first sub-goal (A.3) for the choice of ℓ_n above. Recall that $\ell_n = \mathrm{o}(m)$ and $\ell_n \to \infty$ as $n \to \infty$. Throughout, since we handle $\ell < \ell_n$, we assume ℓ, q, m are positive integers with $m \ge \ell + q$ and $\ell \ge 2$ (if $\ell = 1$ there is nothing to prove).

Lemma A.3 (Sub-goal (A.3))**.** *If Conditions 1 and 2 are satisfied, then sub-goal (A.3) holds, i.e., as $n \to \infty$,*

$$\max_{\ell \in [1:\ell_n]} \left|\ell^{-1}\zeta_{m,\ell} - \zeta_{m,1}\right| = \mathrm{o}(1/m).$$

Proof of sub-goal (A.3). Consider the counting variable

$$N_{\ell,m} = \sum_{i=m-\ell+1}^{m} \mathbb{1}_{i,\ell,m},$$

where

$$\mathbb{1}_{i,\ell,m} = \mathbb{1}\big\{\{Z_i \text{ is in the top-}q \text{ of } Z_{[1:m]}\} \\ \cup \{Z_i \text{ is in the top-}q \text{ of } Z_{[(m-\ell+1):(2m-\ell)]}\}\big\}. \tag{A.10}$$

Define

$$\Xi_{\ell,m} = K_m(Z_{[1:m]})\, K_m(Z_{[(m-\ell+1):(2m-\ell)]}),$$

and write $\mathbb{E}[\Xi; A] = \mathbb{E}[\Xi\, \mathbb{1}_A]$ for a random variable Ξ and an event A. Then,

$$\zeta_{m,\ell} = \mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 0] + \mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 1] + \mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} \geq 2] - \theta^2$$
$$\text{and} \quad \zeta_{m,1} = \mathbb{E}[\Xi_{1,m}; N_{1,m} = 0] + \mathbb{E}[\Xi_{1,m}; N_{1,m} = 1] - \theta^2.$$

We obtain the following decomposition of the left-hand side of (A.3) into three terms:

$$\begin{aligned} |\zeta_{m,\ell} - \ell\zeta_{m,1}| &\leq \left|\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 0] - \ell\, \mathbb{E}[\Xi_{1,m}; N_{1,m} = 0] + (\ell-1)\theta^2\right| \\ &\quad + |\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 1] - \ell\, \mathbb{E}[\Xi_{1,m}; N_{1,m} = 1]| \\ &\quad + |\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} \geq 2]| \\ &:= \mathbb{I}_1 + \mathbb{I}_2 + \mathbb{I}_3 \end{aligned} \tag{A.11}$$

According to Lemmas A.4, A.5 and A.6 below, we have

$$\mathbb{I}_j \lesssim \ell \left(\ell/m^2\right)^{1-2/a} \tag{A.12}$$

for $j \in \{1, 2, 3\}$ and any $a > 2$, where $a_n \lesssim b_n$ if and only if there exists a constant $c > 0$ such that $a_n \leq c b_n$ for all n. Recall the choice of ℓ_n in (A.6). Together with Condition 2, the bounds on $\mathbb{I}_j$ will then imply, as $n \to \infty$ and for sufficiently large a, that

$$\begin{aligned} \max_{\ell \in [1:\ell_n]} \left|\ell^{-1}\zeta_{m,\ell} - \zeta_{m,1}\right| &\lesssim \max_{\ell \in [1:\ell_n]} \left(\ell/m^2\right)^{1-2/a} \\ &\leq n^{2/a-1} + (2^{-1}\ln(n))^{1/2-1/a} m^{3/a-3/2} = \mathrm{o}(1/m). \end{aligned}$$

We conclude that sub-goal (A.3) is attained. □

We now list Lemmas A.4, A.5 and A.6, providing the bounds (A.12). The proofs of these lemmas rely on counting variable inequalities which are developed afterwards in Lemmas A.7 to A.11.

Lemma A.4 (Handling $\mathbb{I}_1$ in (A.11))**.** *If Condition 1 is satisfied, then*

$$\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 0] = \mathbb{P}(N_{\ell,m} = 0)\, \theta^2.$$

In addition,

$$\mathbb{I}_1 \leq \theta^2 \cdot \frac{q^2 \ell(\ell-1)}{(m-1)(m-\ell+1)}.$$

Proof. Let $r = m - \ell$. If $N_{\ell,m} = 0$, the top-q of $Z_{[1:m]}$ is the same as the top-q of $Z_{[1:(m-\ell)]}$ and the top-q of $Z_{[(m-\ell+1):(2m-\ell)]}$ is the same as the one of $Z_{[(m+1):(2m-\ell)]}$. The event $\{N_{\ell,m} = 0\}$ is a function of the random variables $Z_{[(r+1):m]}$ and the q-largest order statistics of $Z_{[1:r]}$ and $Z_{[(m+1):(m+r)]}$. These statements taken together with a slight extension of the properties of the top order statistics of a generalised Pareto sample in Lemma 2.2 imply

$$\begin{aligned}\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 0] &= \mathbb{E}[K_r(Z_{[1:r]})K_r(Z_{[(m+1):(m+r)]}); N_{\ell,m} = 0] \\ &= \mathbb{P}(N_{\ell,m} = 0)\, \theta^2.\end{aligned}$$

Thanks to that equality, we find

$$\begin{aligned}&\left|\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 0] - \ell\, \mathbb{E}[\Xi_{1,m}; N_{1,m} = 0] + (\ell-1)\theta^2\right| \\ &= |\mathbb{P}(N_{\ell,m} = 0) - \ell\, \mathbb{P}(N_{1,m} = 0) + \ell - 1|\, \theta^2 \\ &= |\ell\, \mathbb{P}(N_{1,m} \geq 1) - \mathbb{P}(N_{\ell,m} \geq 1)|\, \theta^2.\end{aligned}$$

Apply Lemma A.10 to conclude. □

In case $N_{\ell,m} = 1$, exactly one of the ℓ indicators $\mathbb{1}_{i,\ell,m}$ in (A.10) for $i \in [(m-\ell+1):m]$ is equal to one while the other $\ell - 1$ indicators are zero. We find

$$\begin{aligned}\{N_{\ell,m} = 1\} &= \bigcup_{i=m-\ell+1}^{m} B_{i,\ell,m} \qquad \text{where} \\ B_{i,\ell,m} &= \{\mathbb{1}_{i,\ell,m} = 1\} \cap \bigcap_{j\in[(m-\ell+1):m]\setminus\{i\}} \{\mathbb{1}_{j,\ell,m} = 0\},\end{aligned} \tag{A.13}$$

and the union over i being disjoint. Note that $B_{i,\ell,m}$ is the event that Z_i is in the top-q of $Z_{[1:m]}$ or $Z_{[(m-\ell+1):(2m-\ell)]}$ while the other variables Z_j for $j \in [(m-\ell+1):m] \setminus \{i\}$ are not. Further, $\{N_{1,m} = 1\} = \{\mathbb{1}_{m,1,m} = 1\}$.

Lemma A.5 (Handling $\mathbb{I}_2$ in (A.11))**.** *If Condition 1 is satisfied, we have*

$$\mathbb{I}_2 \leq 4(5q^2 - q)^{1-2/a} \|K\|_a^2 \ell \left(\frac{(\ell-1)}{(m-\ell+1)^2}\right)^{1-2/a},$$

for all $a > 2$, provided that m is sufficiently large.

Proof. By symmetry and since the union in (A.13) is disjoint,

$$\begin{aligned}\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 1] &= \sum_{i=m-\ell+1}^{m} \mathbb{E}[\Xi_{\ell,m}; B_{i,\ell,m}] \\ &= \ell\, \mathbb{E}[\Xi_{\ell,m}; B_{m,\ell,m}].\end{aligned}$$

On $B_{m,\ell,m}$, the top-q order statistics of $Z_{[(m-\ell+1):(2m-\ell)]}$ are the same as those of $Z_{[m:(2m-\ell)]}$, which implies $K_m(Z_{[(m-\ell+1):(2m-\ell)]}) = K_{m-\ell+1}(Z_{[m:(2m-\ell)]})$ and thus

$$\mathbb{E}[\Xi_{\ell,m}; B_{m,\ell,m}] = \mathbb{E}\left[K_m(Z_{[1:m]})\, K_{m-\ell+1}(Z_{[m:(2m-\ell)]}); B_{m,\ell,m}\right].$$

On the right-hand side, we replace the indicator of the event $B_{m,\ell,m}$ first by $\mathbb{1}_{m,\ell,m}$ and then by $\mathbb{1}_{m,1,m}$, using Lemmas A.11 and A.9, respectively. We find, for $a > 2$, by the triangle inequality and Hölder's inequality,

$$
\begin{aligned}
&\left|\mathbb{E}[\Xi_{\ell,m}; B_{m,\ell,m}] - \mathbb{E}\left[K_m(Z_{[1:m]})\, K_{m-\ell+1}(Z_{[m:(2m-\ell)]})\, \mathbb{1}_{m,1,m}\right]\right| \\
&\le \|K\|_a^2\, \mathbb{E}\left[\left|\mathbb{1}_{B_{m,\ell,m}} - \mathbb{1}_{m,\ell,m}\right|\right]^{1-2/a} + \|K\|_a^2\, \mathbb{E}\left[\left|\mathbb{1}_{m,\ell,m} - \mathbb{1}_{m,1,m}\right|\right]^{1-2/a} \\
&\le \|K\|_a^2 \left[\left(2(2q^2-q)\frac{\ell-1}{(m-1)(m-\ell+1)}\right)^{1-2/a} + \left((q+1)q\frac{\ell-1}{(m+1)m}\right)^{1-2/a}\right] \\
&\le 2(5q^2-q)^{1-2/a}\|K\|_a^2 \left(\frac{\ell-1}{(m-1)(m-\ell+1)}\right)^{1-2/a}.
\end{aligned}
$$

The top-q order statistics of $Z_{[m:(2m-\ell)]}$ are the same as those of $Z_{[m:(2m-1)]}$ unless at least one of the variables Z_j for $j \in [(2m-\ell+1){:}(2m-1)]$ is larger than the q-largest order statistic of $Z_{[m:(2m-\ell)]}$. Let $C_{\ell,m}$ denote the latter event. It follows that, for $a > 2$,

$$
\begin{aligned}
&\Big|\mathbb{E}\left[K_m(Z_{[1:m]})\, K_{m-\ell+1}(Z_{[m:(2m-\ell)]})\, \mathbb{1}_{m,1,m}\right] \\
&\qquad - \mathbb{E}\left[K_m(Z_{[1:m]})\, K_m(Z_{[m:(2m-1)]})\, \mathbb{1}_{m,1,m}\right]\Big| \\
&\le \mathbb{E}\left[\left|K_m(Z_{[1:m]})\right| \left|K_{m-\ell+1}(Z_{[m:(2m-\ell)]}) - K_m(Z_{[m:(2m-1)]})\right| \mathbb{1}_{m,1,m}\right] \\
&\le 2\|K\|_a^2\, \mathbb{E}\left[\mathbb{1}_{m,1,m}; C_{\ell,m}\right]^{1-2/a}.
\end{aligned}
$$

Further, we have

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{m,1,m}; C_{\ell,m}\right] \\
&= \mathbb{P}\Bigg(\left\{Z_m \text{ is in the top-}q \text{ of } Z_{[1:m]} \text{ and/or } Z_{[m:(2m-1)]}\right\} \\
&\quad \cap \bigcup_{j=2m-\ell+1}^{2m-1} \left\{Z_j \text{ is larger than the } q\text{-largest order statistic of } Z_{[m:(2m-\ell)]}\right\}\Bigg).
\end{aligned}
$$

The latter probability is bounded by $\ell - 1$ times the sum

$$
\begin{aligned}
&\mathbb{P}\big(Z_m \text{ is in the top-}q \text{ of } Z_{[1:m]} \text{ and} \\
&\qquad Z_{2m-\ell+1} \text{ is in the top-}q \text{ of } Z_{[(m+1):(2m-\ell+1)]}\big) \\
&+ \mathbb{P}\big(Z_m \text{ and } Z_{2m-\ell+1} \text{ are both in the top-}q \text{ of } Z_{[m:(2m-\ell+1)]}\big)
\end{aligned}
$$

and thus by

$$
(\ell-1)\left(\frac{q}{m}\cdot\frac{q}{m-\ell+1} + \frac{q(q-1)}{(m-\ell+2)(m-\ell+1)}\right) \le (2q^2-q)\frac{\ell-1}{(m-\ell+1)^2}.
$$

Since,

$$
\mathbb{E}\left[K_m(Z_{[1:m]})\, K_m(Z_{[m:(2m-1)]})\, \mathbb{1}_{m,1,m}\right] = \mathbb{E}\left[\Xi_{1,m}; N_{1,m} = 1\right],
$$

a combination of the previous inequalities yields

$$\begin{aligned}
&|\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} = 1] - \ell\,\mathbb{E}[\Xi_{1,m}; N_{1,m} = 1]| \\
&= \ell \left|\mathbb{E}[\Xi_{\ell,m}; B_{m,\ell,m}] - \mathbb{E}\left[K_m(Z_{[1:m]})\,K_m(Z_{[m:(2m-1)]})\,\mathbb{1}_{m,1,m}\right]\right| \\
&\le \ell\|K\|_a^2 \Bigg[2(5q^2-q)^{1-2/a}\left(\frac{\ell-1}{(m-1)(m-\ell+1)}\right)^{1-2/a} \\
&\qquad\qquad + 2(2q^2-q)^{1-2/a}\left(\frac{\ell-1}{(m-\ell+1)^2}\right)^{1-2/a}\Bigg] \\
&\le 4(5q^2-q)^{1-2/a}\|K\|_a^2\ell\left(\frac{(\ell-1)}{(m-\ell+1)^2}\right)^{1-2/a}.
\end{aligned}$$

□

Lemma A.6 (Handling $\mathbb{I}_3$ in (A.11))**.** *If Condition 1 is satisfied, then*

$$\mathbb{I}_3 \le \|K\|_a^2\left(\frac{\ell(\ell-1)}{2}\right)^{1-2/a}\left(\frac{2(2q^2-q)}{(m-1)(m-\ell+1)}\right)^{1-2/a}$$

for all $a > 2$.

Proof. Since the kernel satisfies Condition 1, we have by Hölder's inequality

$$|\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} \ge 2]| \le \mathbb{E}[|\Xi_{\ell,m}|; N_{\ell,m} \ge 2] \le \|K\|_a^2\,\mathbb{P}(N_{\ell,m} \ge 2)^{1-2/a}$$

for $a > 2$. By Lemma A.7, we have

$$\mathbb{P}(N_{\ell,m} \ge 2) \le \frac{\ell(\ell-1)}{2}\,\mathbb{E}[\mathbb{1}_{m-1,\ell,m}\,\mathbb{1}_{m,\ell,m}].$$

In view of Lemma A.8, we conclude

$$|\mathbb{E}[\Xi_{\ell,m}; N_{\ell,m} \ge 2]| \le \|K\|_a^2\left(\frac{\ell(\ell-1)}{2}\right)^{1-2/a}\left(\frac{2(2q^2-q)}{(m-1)(m-\ell+1)}\right)^{1-2/a}. \quad \square$$

We now develop the required counting variable inequalities. In doing so, the following preliminary lemma will be useful a number of times. No originality is claimed.

Lemma A.7. *Consider a counting variable $N = \sum_{i=1}^n \mathbb{1}_i$ with indicator variables $\mathbb{1}_1, \ldots, \mathbb{1}_n$ satisfying $\mathbb{E}[\mathbb{1}_i\,\mathbb{1}_j] \le \varepsilon$ for all $i, j \in [1{:}n]$ and $i \ne j$. Then*

$$\begin{aligned}
2\,\mathbb{P}(N \ge 2) &\le \mathbb{E}(N) - \mathbb{P}(N = 1) \\
&\le 2\left(\mathbb{E}(N) - \mathbb{P}(N \ge 1)\right) \\
&\le n(n-1)\varepsilon.
\end{aligned}$$

Proof. A case-by-case analysis reveals

$$\mathbb{1}\{N \ge 2\} \le \frac{1}{2}(N - \mathbb{1}\{N = 1\}) \le N - \mathbb{1}\{N \ge 1\} \le \frac{1}{2}N(N-1)$$

and thus

$$\mathbb{P}(N \geq 2) \leq \frac{1}{2}\left(\mathbb{E}(N) - \mathbb{P}(N=1)\right) \leq \mathbb{E}(N) - \mathbb{P}(N \geq 1) \leq \frac{1}{2}\left(\mathbb{E}(N^2) - \mathbb{E}(N)\right).$$

But since $\mathbb{1}_i^2 = \mathbb{1}_i$, we have

$$\mathbb{E}(N^2) = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[\mathbb{1}_i \, \mathbb{1}_j] = \sum_{i=1}^n \mathbb{E}[\mathbb{1}_i] + \sum_{i,j \in [1:n], i \neq j} \mathbb{E}[\mathbb{1}_i \, \mathbb{1}_j] \leq \mathbb{E}(N) + n(n-1)\varepsilon,$$

and thus

$$\frac{1}{2}\left(\mathbb{E}(N^2) - \mathbb{E}(N)\right) \leq \frac{n(n-1)}{2}\varepsilon,$$

yielding the stated inequalities. □

Recall the indicator variables $\mathbb{1}_{i,\ell,m}$ in (A.10).

Lemma A.8. *We have*

$$\mathbb{E}[\mathbb{1}_{m-1,\ell,m} \, \mathbb{1}_{m,\ell,m}] \leq \frac{2(2q^2 - q)}{(m-1)(m-\ell+1)}.$$

Proof. By symmetry,

$$\begin{aligned}
&\mathbb{E}[\mathbb{1}_{m-1,\ell,m} \, \mathbb{1}_{m,\ell,m}] \\
&= \mathbb{P}\left(Z_{m-1} \text{ and } Z_m \text{ are both in the top-}q \text{ of } Z_{[1:m]} \text{ or } Z_{[(m-\ell+1):(2m-\ell)]}\right) \\
&\leq 2\,\mathbb{P}\left(Z_{m-1} \text{ and } Z_m \text{ are both in the top-}q \text{ of } Z_{[1:m]}\right) \\
&\quad + 2\,\mathbb{P}\big(Z_{m-1} \text{ is in the top-}q \text{ of } Z_{[1:m]} \\
&\qquad\qquad \text{and } Z_m \text{ is in the top-}q \text{ of } Z_{[(m-\ell+1):(2m-\ell)]}\big).
\end{aligned}$$

The first probability on the right-hand side is equal to $q(q-1)/[m(m-1)]$ while the second one is bounded by

$$\begin{aligned}
&\mathbb{P}\left(Z_{m-1} \text{ is in the top-}q \text{ of } Z_{[1:(m-1)]} \text{ and } Z_m \text{ is in the top-}q \text{ of } Z_{[m:(2m-\ell)]}\right) \\
&\qquad = \frac{q}{m-1} \cdot \frac{q}{m-\ell+1}.
\end{aligned}$$

Add the two bounds and note that $1/m < 1/(m-\ell+1)$ to conclude the proof. □

Lemma A.9. *We have*

$$\mathbb{E}\left[\left|\mathbb{1}_{m,\ell,m} - \mathbb{1}_{m,1,m}\right|\right] \leq (q+1)q\frac{\ell-1}{(m+1)m}.$$

Proof. For events A and B, we have $|\mathbb{1}_A - \mathbb{1}_B| = \mathbb{1}_{A \triangle B}$ where the symmetric difference $A \triangle B = (A \setminus B) \cup (B \setminus A)$ is the event that exactly one of A and B occurs. It follows that $|\mathbb{1}_{m,\ell,m} - \mathbb{1}_{m,1,m}|$ is the indicator of the event that exactly one of the two events

$$\{Z_m \text{ is in the top-}q \text{ of } Z_{[1:m]} \text{ or } Z_m \text{ is in the top-}q \text{ of } Z_{[(m-\ell+1):(2m-\ell)]}\} \text{ and}$$
$$\{Z_m \text{ is in the top-}q \text{ of } Z_{[1:m]} \text{ or } Z_m \text{ is in the top-}q \text{ of } Z_{[m:(2m-1)]}\}$$

occurs. In that case, Z_m is either in the top-q of $Z_{[(m-\ell+1):(2m-\ell)]}$ but not in the top-q of $Z_{[m:(2m-1)]}$ or the other way around, and thus

$$
\begin{aligned}
&\mathbb{E}\left[\left|\mathbb{1}_{m,\ell,m} - \mathbb{1}_{m,1,m}\right|\right] \\
&\le \mathbb{P}\left(Z_m \text{ is in the top-}q \text{ of } Z_{[(m-\ell+1):(2m-\ell)]} \text{ but not in that of } Z_{[m:(2m-1)]}\right) \\
&+ \mathbb{P}\left(Z_m \text{ is in the top-}q \text{ of } Z_{[m:(2m-1)]} \text{ but not in that of } Z_{[(m-\ell+1):(2m-\ell)]}\right) \\
&= 2\,\mathbb{P}\left(Z_m \text{ is in the top-}q \text{ of } Z_{[m:(2m-1)]} \text{ but not in that of } Z_{[(m-\ell+1):(2m-\ell)]}\right).
\end{aligned}
$$

For the probability of the latter event, note that there always needs to be at least one of the $\ell - 1$ random variables of $Z_{[(m-\ell+1):(m-1)]}$ larger than Z_m for Z_m not to be in the top-q of $Z_{[(m-\ell+1):(2m-\ell)]}$. Thus, by the union bound, the probability is bounded by $(\ell - 1)$ times the probability that Z_{m-1} and Z_m are both in the top-$(q + 1)$ of $Z_{[(m-1):(2m-1)]}$ while Z_{m-1} is still larger than Z_m. It follows that

$$
|\mathbb{E}[\mathbb{1}_{m,\ell,m} - \mathbb{1}_{m,1,m}]| \le 2(\ell-1)\cdot\frac{(q+1)q}{2(m+1)m} = (q+1)q\frac{\ell-1}{(m+1)m}. \qquad \square
$$

Lemma A.10. *We have*

$$
|\mathbb{P}(N_{\ell,m} \ge 1) - \ell\,\mathbb{P}(N_{1,m} = 1)| \le 3q^2 \cdot \frac{\ell(\ell-1)}{(m-1)(m-\ell+1)}.
$$

Proof. In view of Lemma A.7, we have

$$
|\mathbb{P}(N_{\ell,m} \ge 1) - \ell\,\mathbb{E}[\mathbb{1}_{m,\ell,m}]| \le \frac{\ell(\ell-1)}{2}\,\mathbb{E}[\mathbb{1}_{m-1,\ell,m}\,\mathbb{1}_{m,\ell,m}]
$$

and thus, since $N_{1,m} = \mathbb{1}_{m,1,m}$,

$$
\begin{aligned}
|\mathbb{P}(N_{\ell,m} \ge 1) - \ell\,\mathbb{P}(N_{1,m} = 1)| \le{}& \frac{\ell(\ell-1)}{2}\,\mathbb{E}[\mathbb{1}_{m-1,\ell,m}\,\mathbb{1}_{m,\ell,m}] \\
&+ \ell\,\mathbb{E}\left[\left|\mathbb{1}_{m,\ell,m} - \mathbb{1}_{m,1,m}\right|\right].
\end{aligned}
$$

The two expectations on the right-hand side were treated in Lemmas A.8 and A.9, respectively. Combine the bounds to obtain that

$$
\begin{aligned}
&|\mathbb{P}(N_{\ell,m} \ge 1) - \ell\,\mathbb{P}(N_{1,m} = 1)| \\
&\le \frac{\ell(\ell-1)}{2}\cdot\frac{2(2q^2-q)}{(m-1)(m-\ell+1)} + \ell\cdot(q+1)q\frac{\ell-1}{(m+1)m} \\
&\le \frac{3q^2\ell(\ell-1)}{(m-1)(m-\ell+1)}.
\end{aligned}
$$

$\square$

Lemma A.11. *We have*

$$
\mathbb{E}\left[\left|\mathbb{1}_{B_{m,\ell,m}} - \mathbb{1}_{m,\ell,m}\right|\right] \le 2(2q^2-q)\frac{\ell-1}{(m-1)(m-\ell+1)}.
$$

Proof. We have $\mathbb{1}_{m,\ell,m} \geq \mathbb{1}_{B_{m,\ell,m}}$ while the difference between the two indicators is equal to the indicator of the event that $\mathbb{1}_{m,\ell,m} = 1$ holds together with $\mathbb{1}_{j,\ell,m} = 1$ for some $j \in [(m-\ell+1){:}(m-1)]$. By the union bound, we obtain

$$\mathbb{E}\left[\left|\mathbb{1}_{B_{m,\ell,m}} - \mathbb{1}_{m,\ell,m}\right|\right] \leq (\ell-1)\,\mathbb{E}[\mathbb{1}_{m-1,\ell,m}\,\mathbb{1}_{m,\ell,m}].$$

Apply Lemma A.8 to conclude. □

Appendix B: Proof of Proposition 2.4

Proof of Proposition 2.4. The proof is divided in two parts. The first part will deal with the expression for the asymptotic variance in (2.16) and the second part with the asymptotic normality of the Hájek projection in (2.17).

Part 1: asymptotic variance. Write $\overline{K}_m := K_m - u(\gamma)$. We start by calculating the limit

$$\lim_{r\to\infty} r\,\zeta_{r+1,1}, \quad \text{where} \quad \zeta_{r+1,1} = \mathbb{E}\left[\overline{K}_{r+1}(Z_1,\ldots,Z_r,Z)\cdot\overline{K}_{r+1}(Z_1',\ldots,Z_r',Z)\right]$$

and $Z, Z_1, Z_1', Z_2, Z_2', \ldots$ are independent $\mathrm{GP}(\gamma)$ random variables. The expectation defining $\zeta_{r+1,1}$ is finite by Condition 1 and the Cauchy–Schwarz inequality. We will prove that

$$\lim_{r\to\infty} r\,\zeta_{r+1,1} = \sigma_{\gamma,K}^2 = \int_0^\infty \left(\mathbb{E}\left[\overline{K}_{q+1}\big((Z_{q-j:q-1})_{j=1}^q, h_\gamma(S_q/x)\big)\right]\right)^2 \mathrm{d}x < \infty, \tag{B.1}$$

where $Z_{0:q-1} = 0$ and $Z_{1:q-1} \leq \ldots \leq Z_{q-1:q-1}$ are the order statistics of an independent $\mathrm{GP}(\gamma)$ random sample $Z_1,\ldots,Z_{q-1}$, while $S_q = E_1 + \cdots + E_q$ with $E_1,\ldots,E_q$ an independent unit-exponential random sample, independent of $Z_1,\ldots,Z_{q-1}$.

To prove (B.1), we will show that we can rewrite $r\,\zeta_{r+1,1}$ as

$$r\,\zeta_{r+1,1} = \int_0^\infty \left(\mathbb{E}\left[\overline{K}_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x)\big)\cdot\mathbb{1}\{rU_{q:r} > x\}\right]\right)^2 \mathrm{d}x, \tag{B.2}$$

where $Y_{0:q-1} := 1$, $Y_{1:q-1} \leq \ldots \leq Y_{q-1:q-1}$ are the order statistics of a standard Pareto random sample of size $q-1$ and $U_{1:r} \leq \ldots \leq U_{r:r}$ are the order statistics of a uniform$(0,1)$ random sample of size r, the two samples being independent. Since $rU_{q:r}$ converges in distribution to S_q, the remainder of the proof then consists in justifying why the limit can be interchanged with the expectation and the integral. We will do this in two steps: first we will show that the expectation inside the integral in (B.2) converges, i.e., for all $x > 0$, we have the convergence

$$\begin{aligned}\lim_{r\to\infty} \mathbb{E}&\left[\overline{K}_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x)\big)\cdot\mathbb{1}\{rU_{q:r} > x\}\right] \\ &= \mathbb{E}\left[\overline{K}_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(S_q/x)\big)\cdot\mathbb{1}\{S_q > x\}\right],\end{aligned} \tag{B.3}$$

where S_q is as in (B.1). Second, we will apply Lebesque's dominated convergence theorem to show that the integral itself converges.

We start with the proof of (B.2). Condition on Z to get

$$\begin{aligned} r\,\zeta_{r+1,1} &= \int_0^\infty \mathbb{E}\left[\overline{K}_{r+1}(Z_1,\ldots,Z_r,z)\cdot\overline{K}_{r+1}(Z_1',\ldots,Z_r',z)\right]\cdot r\cdot \mathrm{d}\,\mathbb{P}(Z\le z) \\ &= \int_0^\infty \left(\mathbb{E}\left[\overline{K}_{r+1}(Z_1,\ldots,Z_r,z)\right]\right)^2\cdot r\cdot \mathrm{d}\,\mathbb{P}(Z\le z) \\ &= \int_0^\infty \left(\mathbb{E}\left[\overline{K}_{q+1}\big((Z_{r-j+1:r})_{j=1}^q,z\big)\right]\right)^2\cdot r\cdot \mathrm{d}\,\mathbb{P}(Z\le z), \end{aligned}$$

in terms of the ascending order statistics $Z_{1:r}\le\ldots\le Z_{r:r}$ of $Z_1,\ldots,Z_r$. Here we used the fact that the kernel K_m only depends on the top-q values of its argument. Note that if Y is a standard Pareto random variable, then the distribution of $h_\gamma(Y)$ is $\mathrm{GP}(\gamma)$. Let $Y_{1:r}\le\ldots\le Y_{r:r}$ be the ascending order statistics of a standard Pareto random sample $Y_1,\ldots,Y_r$. It follows that

$$\begin{aligned} r\,\zeta_{r+1,1} &= \int_1^\infty \left(\mathbb{E}\left[\overline{K}_{q+1}\big((h_\gamma(Y_{r-j+1:r}))_{j=1}^q,h_\gamma(y)\big)\right]\right)^2\cdot r\frac{\mathrm{d}y}{y^2} \\ &= \int_{1/r}^\infty \left(\mathbb{E}\left[\overline{K}_{q+1}\big((h_\gamma(Y_{r-j+1:r}))_{j=1}^q,h_\gamma(rt)\big)\right]\right)^2\frac{\mathrm{d}t}{t^2}, \end{aligned}$$

where we have substituted $y=rt$ in the last step. The vector of top-q standard Pareto order statistics satisfies the distributional representation

$$(Y_{r-j+1:r})_{j=1}^q \stackrel{d}{=} (Y_{q-j:q-1}/U_{q:r})_{j=1}^q$$

where $Y_{0:q-1}:=1$ and where $U_{1:r}\le\ldots\le U_{r:r}$ denote the ascending order statistics of an independent uniform$(0,1)$ sample $U_1,\ldots,U_r$, independent of $Y_1,\ldots,Y_r$; note that $1/U_{q:r}$ has the same distribution as $Y_{r-q+1:r}$. The function h_γ satisfies the functional equation

$$\forall x,y>0,\qquad h_\gamma(y/x)=\frac{h_\gamma(y)-h_\gamma(x)}{x^\gamma}.$$

By location-scale invariance of the kernel,

$$\begin{aligned} &\overline{K}_{q+1}\big((h_\gamma(Y_{q-j:q-1}/U_{q:r}))_{j=1}^q,h_\gamma(rt)\big) \\ &= \overline{K}_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q,U_{q:r}^\gamma h_\gamma(rt)+h_\gamma(U_{q:r})\big) \\ &= \overline{K}_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q,h_\gamma(rU_{q:r}t)\big). \end{aligned}$$

Following the substitution $x=1/t$, we obtain

$$\begin{aligned} r\,\zeta_{r+1,1} &= \int_{1/r}^\infty \left(\mathbb{E}\left[\overline{K}_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q,h_\gamma(rU_{q:r}t)\big)\right]\right)^2\frac{\mathrm{d}t}{t^2} \\ &= \int_0^r \left(\mathbb{E}\left[\overline{K}_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q,h_\gamma(rU_{q:r}/x)\big)\right]\right)^2\mathrm{d}x. \end{aligned}$$

By independence and Fubini's theorem, the expectation inside the integral is equal to

$$\begin{aligned}
&\mathbb{E}\left[\overline{K}_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x)\big)\right] \\
&\qquad = \int_0^1 \mathbb{E}\left[\overline{K}_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(ru/x)\big)\right] \,\mathrm{d}\,\mathbb{P}(U_{q:r} \le u).
\end{aligned}$$

The expectation inside the integral on the right-hand side is zero as soon as $ru/x \le 1$, because in that case

$$\begin{aligned}
&\mathbb{E}\left[\overline{K}_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(ru/x)\big)\right] \\
&\qquad = \mathbb{E}\left[\overline{K}_{q}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q\big)\right] = \mathbb{E}\left[\overline{K}_{q}\big((Z_{q-j:q-1})_{j=1}^q\big)\right] = 0.
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\overline{K}_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x)\big)\right] \\
&\qquad = \mathbb{E}\left[\overline{K}_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x)\big) \cdot \mathbb{1}\{rU_{q:r} > x\}\right].
\end{aligned}$$

Equation (B.2) follows as the integral is equal to 0 for values of x higher than r.

Next, we show (B.3) and eventually (B.1). The well-known representation of uniform order statistics in terms of partial sums of unit-exponential random variables, i.e., $U_{q:r} \stackrel{d}{=} S_q/S_{r+1}$ where, for any integer a, $S_a = E_1 + \cdots + E_a$ is the sum of independent standard exponential random variables independent of $Y_1, \ldots, Y_{q-1}$, implies the weak convergence

$$rU_{q:r} \xrightarrow{d} S_q = E_1 + \cdots + E_q, \qquad r \to \infty. \tag{B.4}$$

The question is whether in (B.3), we can switch the limit and the expectation.

Under a suitable Skorokhod construction, we can assume that (B.4) holds almost surely. By continuity of the kernel, the integrand inside the expectation on the left-hand side then converges almost surely to the integrand on the right-hand side. To show uniform integrability in (B.3) and eventually the validity of switching the limit and integral in (B.1), it is sufficient to show that the expectation of the square is bounded by an integrable function $c(x)$, i.e.,

$$\limsup_{r\to\infty} \mathbb{E}\left[\overline{K}_{q+1}^2\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x)\big) \cdot \mathbb{1}\{rU_{q:r} > x\}\right] < c(x) < \infty,$$

where $c(x)$ is a function on $[0, +\infty)$ such that $\int_0^{+\infty} c(x)\mathrm{d}x < \infty$. Then not only the relation (B.3) follows immediately, but also the relation (B.1) follows from the Lebesgue dominance convergence theorem.

To do so, we rely on a change-of-measure, replacing $rU_{q:r}/x$ on the event $\{rU_{q:r} > x\}$ by a standard Pareto random variable, and compensating by the likelihood ratio. For $0 < x < r$, let $p_{r,x}$ be the probability density function of $rU_{q:r}/x$ and let Y be a standard Pareto random variable, independent of $Y_1, \ldots, Y_{q-1}$. We have

$$\begin{aligned}
&\mathbb{E}\left[\overline{K}^2_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(rU_{q:r}/x)\big) \cdot \mathbb{1}\{rU_{q:r} > x\}\right] \\
&= \int_1^\infty \mathbb{E}\left[\overline{K}^2_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(y)\big)\right] \cdot p_{r,x}(y)\,\mathrm{d}y \\
&= \int_1^\infty \mathbb{E}\left[\overline{K}^2_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(y)\big)\right] \cdot p_{r,x}(y) \cdot y^2\,\frac{\mathrm{d}y}{y^2} \\
&= \mathbb{E}\left[\overline{K}^2_{q+1}\big((h_\gamma(Y_{q-j:q-1}))_{j=1}^q, h_\gamma(Y)\big) \cdot p_{r,x}(Y) \cdot Y^2\right] \\
&= \mathbb{E}\left[\overline{K}^2_{q}\big(h_\gamma(Y_1), \ldots, h_\gamma(Y_{q-1}), h_\gamma(Y)\big) \cdot p_{r,x}(Y) \cdot Y^{2-\epsilon} \cdot Y^{\epsilon}\right],
\end{aligned}$$

for some $0 < \epsilon < 1/2$.

Recall that the distribution of $(h_\gamma(Y_1), \ldots, h_\gamma(Y_{q-1}), h_\gamma(Y))$ is equal to the one of an independent $\mathrm{GP}(\gamma)$ random sample of size q. In view of Condition 1, and the fact that $\mathbb{E}[Y^{2\epsilon}] < \infty$, it is sufficient to show that

$$\sup_{r>2q}\sup_{y\geq 1} p_{r,x}(y) \cdot y^{2-\epsilon} \leq \tilde{c}(x),$$

for some finite function $\tilde{c}(x)$ such that $\int_0^{+\infty} \tilde{c}(x)\mathrm{d}x < \infty$. Since the distribution of $U_{q:r}$ is $\mathrm{Beta}(q, r+1-q)$, we have, for $y \in [1, \infty)$,

$$\begin{aligned}
p_{r,x}(y) &= \frac{\mathrm{d}}{\mathrm{d}y}\,\mathbb{P}(rU_{q:r}/x \leq y) \\
&= \frac{\mathrm{d}}{\mathrm{d}y}\,\mathbb{P}(U_{q:r} \leq xy/r) \\
&= (x/r) \cdot \frac{\Gamma(r+1)}{\Gamma(q)\Gamma(r+1-q)} \cdot (xy/r)^{q-1} \cdot (1-xy/r)^{r-q} \cdot \mathbb{1}\{xy < r\} \\
&= \frac{x^q}{(q-1)!} \cdot \frac{r!}{(r-q)!r^q} \cdot y^{q-1} \cdot (1-xy/r)^{r-q} \cdot \mathbb{1}\{xy < r\}. \qquad \text{(B.5)}
\end{aligned}$$

As $1 - h \leq e^{-h}$ for real h, we get, for $r > 2q$, the inequalities

$$p_{r,x}(y) \leq \frac{x^q}{(q-1)!} \cdot y^{q-1} \cdot e^{-(r-q)xy/r} \leq \frac{x^q}{(q-1)!} \cdot y^{q-1} \cdot e^{-xy/2}.$$

Hence,

$$p_{r,x}(y) \cdot y^{2-\epsilon} \leq \frac{x^q}{(q-1)!} \cdot y^{q+1-\epsilon} \cdot e^{-xy/2}, \quad \text{for all } r \geq q, \text{ all } x > 0 \text{ and all } y \geq 1.$$

With straightforward calculation, we get that the upperbound is maximized at $y = 1 \vee 2(q+1-\epsilon)/x$. It follows that we can choose

$$\tilde{c}(x) = \begin{cases} \dfrac{(2(q+1-\epsilon))^{q+1-\epsilon}}{(q-1)!e^{q+1-\epsilon}} \cdot x^{-1+\epsilon} & \text{if } 0 < x \leq 2(q+1-\epsilon), \\[2ex] \dfrac{x^q}{(q-1)!} \cdot e^{-x/2} & \text{if } x > 2(q+1-\epsilon), \end{cases}$$

which satisfies $\int_0^\infty \tilde{c}(x)\,\mathrm{d}x < \infty$. This concludes the proof of (B.3) and (B.1).

Part 2: asymptotic normality. It remains to show that the Hájek projection $\hat{U}^m_{n,Z}$ in (2.11) of $U^m_{n,Z} - \theta$ is asymptotically normal. Note that $\hat{U}^m_{n,Z}$ is a centered row sum of a triangular array of row-wise i.i.d. random variables, with total variance $\zeta_{m,1} m^2/n$. If Lyapunov's condition is met, following the central limit theorem, we have

$$\frac{k}{\sqrt{n\zeta_{m,1}}}\hat{U}^m_{n,Z} \xrightarrow{d} N(0,1), \qquad n \to \infty.$$

Therefore, we need to show that for some $\delta > 0$ we have

$$\lim_{m\to\infty} \frac{1}{\left(\sqrt{n\zeta_{m,1}}\right)^{2+\delta}} \cdot n\,\mathbb{E}\left[\left|\hat{K}_m(Z)\right|^{2+\delta}\right] = 0. \tag{B.6}$$

First, by Jensen's inequality,

$$\left|\hat{K}_m(z)\right|^{2+\delta} \leq \mathbb{E}\left[\left|\overline{K}_m(Z_1,\ldots,Z_{m-1},z)\right|^{2+\delta}\right].$$

Second, by Fubini's theorem, location-scale invariance of the kernel and the stability property of GP order statistics,

$$\mathbb{E}\left[\left|\hat{K}_m(Z)\right|^{2+\delta}\right] \leq \mathbb{E}\left[\left|\overline{K}_m(Z_1,\ldots,Z_m)\right|^{2+\delta}\right] = \mathbb{E}\left[\left|\overline{K}_q(Z_1,\ldots,Z_q)\right|^{2+\delta}\right],$$

which implies that the term $\mathbb{E}[|\hat{K}_m(Z)|^{2+\delta}]$ is uniformly bounded over all m.

Since $\lim_{m\to\infty} m\zeta_{m,1} = \sigma_K^2(\gamma) > 0$, the relation (B.6) holds provided we can choose $\delta > 0$ such that

$$\lim_{n\to\infty}\left(\frac{m}{n}\right)^{(2+\delta)/2} n = 0.$$

From Condition 2, this can be achieved by choosing $\delta > 2/\epsilon$. ☐

Appendix C: Proof of Proposition 3.2

We start by listing two preliminary lemmas.

Lemma C.1. *(Potter's bounds). Let $f \in RV_\alpha$. For all $\delta_1, \delta_2 > 0$ there exists $t_0 = t_0(\delta_1,\delta_2)$ such that for any positive t and x satisfying $t \geq t_0$ and $tx \geq t_0$,*

$$(1-\delta_1)x^\alpha \min\left(x^{\delta_2}, x^{-\delta_2}\right) \leq \frac{f(tx)}{f(t)} \leq (1+\delta_1)x^\alpha \max\left(x^{\delta_2}, x^{-\delta_2}\right).$$

Lemma C.2. *Let $Y^*_{1:m} \leq Y^*_{2:m} \leq \ldots \leq Y^*_{m:m}$ be the order statistics of an independent random sample of Pareto(1) random variables. For integer $q \geq 1$ and for a real number $\nu < q$, the sequence $\{(Y^*_{m-q+1:m}/m)^\nu : m \geq m_0\}$ is uniformly integrable for some m_0 depending on q and ν.*

Proof. Note that $Y^*_{m-q+1:m} \stackrel{d}{=} \frac{S_{m+1}}{S_q}$, where $S_l = \sum_{i=1}^{l} E_i$ for independent unit-exponential random variables E_i. Since it is only the distribution of $Y^*_{m-q+1:m}$ that matters, we can change probability spaces to ensure that the equality in distribution is in fact an equality between random variables.

We first deal with $(Y^*_{m-q+1:m}/m)^\nu$ for $\nu \le 0$. We have

$$(Y^*_{m-q+1:m}/m)^\nu \le \left(\frac{S_q}{(m+1)^{-1}\sum_{j=1}^{m+1} E_j} \right)^{-\nu}.$$

Note that $H_m = (m+1)^{-1}\sum_{i=j}^{m+1} E_j$ is distributed as a Gamma$(m+1, m+1)$ random variable with value $m+1$ for both the shape and rate parameter. By Cauchy's inequality,

$$\mathbb{E}\left[(Y^*_{m-q+1:m}/m)^{2\nu}\right] \le \sqrt{\mathbb{E}[H_m^{2\nu}]}\sqrt{\mathbb{E}[S_q^{-2\nu}]}.$$

We have $\mathbb{E}[S_q^{-2\nu}] < \infty$ for all $\nu \le 0$ while by direct calculation, provided $m > -2\nu - 1$,

$$\mathbb{E}[H_m^{2\nu}] = \frac{(m+1)^{-2\nu}\Gamma(m+2\nu+1)}{\Gamma(m+1)} \to 1 \text{ as } m \to \infty.$$

Therefore, there exists $m_0(\nu)$ such that the sequence $\{(Y^*_{m-q+1:m}/m)^\nu, m \ge m_0(\nu)\}$ is bounded in $\mathcal{L}^2$ and thus uniformly integrable.

Next we consider $(Y^*_{m-q+1:m}/m)^\nu$ for $0 < \nu < q$. Define $p_0 = (3\nu + q)/4$ and note that $\nu < p_0 < (\nu + q)/2$. Also, define $p_1 = (\nu + q)/(2p_0)$, which satisfies $p_1 > 1$, and define $z > 1$ by $1/z + 1/p_1 = 1$. For sufficiently large integer m, we find, using Hölder's inequality and the inequality $1/m \le 2/(m+1)$, that

$$\begin{aligned}\mathbb{E}\left[(Y^*_{m-q+1:m}/m)^{p_0}\right] &\le 2^{p_0}\,\mathbb{E}\left[(H_m/S_q)^{p_0}\right] \\ &\le 2^{p_0}\left(\mathbb{E}[H_m^{zp_0}]\right)^{1/z}\left(\mathbb{E}[S_q^{-p_1 p_0}]\right)^{1/p_1}.\end{aligned}$$

From $p_0 p_1 = (\nu + q)/2 < q$ it follows that $\mathbb{E}[S_q^{-p_0 p_1}] < \infty$. Similar to the case where $\nu < 0$, for any $a > 0$ we have $\sup_{m \ge 1}\mathbb{E}[H_m^a] < \infty$. It follows that

$$\sup_{m \ge q}\mathbb{E}\left[(Y^*_{m-q+1:m}/m)^{p_0}\right] < \infty.$$

Since $p_0 > \nu > 0$, the sequence $\{(Y^*_{m-q+1:m}/m)^\nu, m \ge q\}$ is uniformly integrable. □

Proof of Proposition 3.2. Recall from the discussion in Section 3 that,

$$\begin{aligned}\sqrt{k}\left(U_n^m - U_{n,Z}^m\right) &= \sqrt{k}\binom{n}{m}^{-1}\sum_{I\subset[n],|I|=m}\left(K_m\left(X_I\right) - K_m\left(Z_I\right)\right) \\ &= \sqrt{k}A(m)\binom{n}{m}^{-1}\sum_{I\subset[n],|I|=m}\xi_m^I,\end{aligned}$$

where $\xi_m^I = K_m(X_I) - K_m(Z_I)$. In what follows we write ξ_m to indicate a generic ξ_m^I where the index set is not important. Denote

$$U_{n,\xi}^m = \binom{n}{m}^{-1} \sum_{I \subset [n], |I|=m} \xi_m^I.$$

We have

$$U_{n,\xi}^m \xrightarrow{p} B_K(\rho, \gamma),$$

if

$$\lim_{m\to\infty} \mathbb{E}\,\xi_m = B_K(\rho,\gamma) \qquad \text{and} \sum_{\ell=1}^m p_{n,m}(\ell)\Psi_{m,\ell} \to 0, \qquad n \to \infty,$$

where $\Psi_{m,\ell} = \operatorname{cov}\left(\xi_m^I, \xi_m^{I'}\right)$ and I, I' are two index sets with exactly ℓ common elements.

We start by proving that $\lim_{m\to\infty} \mathbb{E}\,\xi_m = B_K(\rho,\gamma)$. By exploiting the scale and location invariance of K and Lemma 2.2 for the $\mathrm{GP}(\gamma)$ distributed random variables, ξ_m^I is equal in distribution to

$$\xi_m := \frac{K\left(\{h(Y_{q-j:q-1}; Y^*_{m-q+1:m})\}_{j=1}^q\right) - K\left(\{h_\gamma(Y_{q-j:q-1})\}_{j=1}^q\right)}{A(m)},$$

where $Y_{0:q-1} = 1$, $Y_{q-1:q-1} \ge Y_{q-2:q-1} \ge \ldots \ge Y_{1:q-1}$ are Pareto(1) order statistics independent of $Y^*_{m-q+1:m}$ and $h(x;t)$ is the function defined as

$$h(x;t) = \frac{U(tx) - U(t)}{t\,U'(t)}.$$

By Condition 5 the function K is differentiable and all q first-order partial derivatives $\dot{K}_1, \ldots, \dot{K}_q$ are continuous. The mean value theorem for several variables implies that there exists a point $\hat{Z}_m$ on the line segment between the points $\{h(Y_{q-j:q-1}; Y^*_{m-q+1:m})\}_{j=1}^q$ and $\{Z_{q-j:q-1}\}_{j=1}^q = \{h_\gamma(Y_{q-j:q-1})\}_{j=1}^q$ such that as $m \to \infty$,

$$\begin{aligned} \xi_m &= \frac{A(Y^*_{m-q+1:m})}{A(m)} \sum_{j=1}^q \dot{K}_j(\hat{Z}_m) \frac{h(Y_{q-j:q-1}; Y^*_{m-q+1:m}) - h_\gamma(Y_{q-j:q-1})}{A(Y^*_{m-q+1:m})} \\ &\to S_q^{-\rho} \sum_{j=1}^q \dot{K}_j\left(\{h_\gamma(Y_{q-j:q-1})\}_{j=1}^q\right) H_{\gamma,\rho}(Y_{q-j:q-1}) \end{aligned} \tag{C.1}$$

almost surely, where S_q is a properly constructed Erlang(q) random variable independent of $Y_{q-1:q-1} \ge Y_{q-2:q-1} \ge \ldots \ge Y_{1:q-1}$. To see this, note that under suitable Skorokhod construction, $m^{-1}Y^*_{m-q+1:m} \to S_q^{-1}$ almost surely. In addition, $A(mx)/A(m) \to x^\rho$ as $m \to \infty$. Therefore, it follows by the extended continuous mapping theorem that $A(Y^*_{m-q+1:m})/A(m) \to S_q^{-\rho}$ as $m \to \infty$.

In view of (C.1), we show that the sequence $\{\xi_m\}_{m\geq m_0}$ is uniformly integrable for sufficiently large integer m_0. In particular, we will show that

$$\sup_{m\geq m_0} \mathbb{E}[|\xi_m|^{1+\epsilon/2}] < \infty$$

for some positive integer m_0, with $\epsilon > 0$ as given in Condition 6. Write $\eta = 1+\epsilon/2$. We deal with ξ_m on the event

$$D_m = \{Y^*_{m-q+1:m} > t_0\}$$

and its complement D^c_m for $t_0 > 1$ to be defined later.

By independence of $Y^*_{m-q+1:m}$ and $Y_1, \ldots, Y_{q-1}$, we have, by Cauchy's inequality, for any $t_0 > 1$,

$$\begin{aligned}
\mathbb{E}(|\xi_m|^\eta; D^c_m) &\leq \frac{2^\eta\, \mathbb{E}[|K_m(X_I)|^\eta + |K_m(Z_I)|^\eta; D^c_m]}{|A(m)|^\eta} \\
&\leq \frac{2^\eta}{|A(m)|^\eta} \left\{ [\mathbb{E}(|K_m(X_I)|^{2\eta})]^{1/2} + [\mathbb{E}(|K_m(Z_I)|^{2\eta})]^{1/2} \right\} [\mathbb{P}(D^c_m)]^{1/2} \to 0,
\end{aligned}$$

as $m \to \infty$. To obtain the final step, we use the following three facts: $\mathbb{E}(|K_m(Z_I)|^{2\eta})$ is finite and does not depend on m, $\sup_{m\geq q} \mathbb{E}(|K_m(X_I)|^{2\eta}) = O(m^a)$ for some $a > 0$ by Condition 6 and $\mathbb{P}(D^c_m)$ tends to zero much faster than any power of $|A(m)|$ (exponentially versus polynomially).

For the remaining term $|\xi_m|^\eta\, \mathbb{I}_{D_m}$, consider again the expansion (C.1) of ξ_m by the mean value theorem with

$$\hat{Z}_m = (1-\hat{a})\{h(Y_{q-j:q-1}; Y^*_{m-q+1:m})\}_{j=1}^q + \hat{a}\{h_\gamma(Y_{q-j:q-1})\}_{j=1}^q$$

for some random $0 < \hat{a} < 1$. Fix $q = 3$ and abbreviate $\left(\hat{Z}_{3-j:2}\right)_{j=1}^3 := \hat{Z}_m$, where $\hat{Z}_{2:2} \geq \hat{Z}_{1:2} \geq \hat{Z}_{0:2} = 0$. Fix $\delta > 0$, to be determined later. Note that since $|A|$ is regularly varying, we can use the inequality in Lemma C.1 to obtain that there exists some $t_0(\delta) > 0$ such that

$$|\xi_m|^\eta\, \mathbb{I}_{D_m} \leq (1+\delta) \left(\frac{Y^*_{m-2:m}}{m}\right)^{\eta(\rho\pm\delta)} \left| \sum_{j=1}^3 \dot{K}_j(\hat{Z}_m) \frac{\tilde{h}(Y_{3-j:2})}{A(Y^*_{m-2:m})} \right|^\eta$$

where $(x)^{\pm\delta}$ is short-hand for $\max\{x^\delta, x^{-\delta}\}$ and where $\tilde{h}(Y_{i:2}) = h(Y_{i:2}; Y^*_{m-2:m}) - h_\gamma(Y_{i:2})$. By exploiting the derivative structure, we have

$$\begin{aligned}
&\sum_{j=1}^3 \dot{K}_j(\hat{Z}_m) \tilde{h}(Y_{3-j:2}) \\
&= \dot{K}_1(\hat{Z}_m) \left(\tilde{h}(Y_{2:2}) - \tilde{h}(Y_{1:2})\right) + \dot{K}_3(\hat{Z}_m) \left(\tilde{h}(Y_{0:2}) - \tilde{h}(Y_{1:2})\right) \\
&= g'\left(\ln\left(\tfrac{\hat{Z}_{2:2}-\hat{Z}_{1:2}}{\hat{Z}_{1:2}-\hat{Z}_{0:2}}\right)\right) \left(\frac{\tilde{h}(Y_{2:2}) - \tilde{h}(Y_{1:2})}{\hat{Z}_{2:2}-\hat{Z}_{1:2}} + \frac{\tilde{h}(Y_{0:2}) - \tilde{h}(Y_{1:2})}{\hat{Z}_{1:2}-\hat{Z}_{0:2}}\right).
\end{aligned}$$

It follows that

$$|\xi_m|^\eta \mathbb{I}_{D_m} \le (1+\delta)\left(\frac{Y^*_{m-2:m}}{m}\right)^{\eta(\rho\pm\delta)} \left|g'\left(\ln\left(\frac{\hat{Z}_{2:2}-\hat{Z}_{1:2}}{\hat{Z}_{1:2}-\hat{Z}_{0:2}}\right)\right)\right|^\eta \cdot \left|\sum_{j=1}^{2} \frac{\tilde{h}(Y_{3-j:2})-\tilde{h}(Y_{2-j:2})}{\left(\hat{Z}_{3-j:2}-\hat{Z}_{2-j:2}\right)A(Y^*_{m-2:m})}\right|^\eta .$$

By Lemma C.3 below, any positive power of

$$\left|\sum_{j=1}^{2} \frac{\tilde{h}(Y_{3-j:2})-\tilde{h}(Y_{2-j:2})}{\left(\hat{Z}_{3-j:2}-\hat{Z}_{2-j:2}\right)A(Y^*_{m-2:m})}\right|$$

can be bounded by an integrable random variable not depending on m, provided $Y^*_{m-2:m} > t_0(\delta,\delta_1,\delta_2,\gamma)$. Thus, to prove $\sup_{m>m_0}\mathbb{E}\,|\xi_m|^\eta \mathbb{I}_{D_m} < \infty$, by Hölder's inequality, it suffices to prove that

$$\sup_{m>m_0} \mathbb{E}\left[\left|g'\left(\ln\left(\frac{\hat{Z}_{2:2}-\hat{Z}_{1:2}}{\hat{Z}_{1:2}-\hat{Z}_{0:2}}\right)\right)\right|^{r\eta}\left(\frac{Y^*_{m-2:m}}{m}\right)^{r\eta(\rho\pm\delta)}\right] < \infty$$

for some $r > 1$. In addition, Lemma C.2 shows that $\left(Y^*_{m-2:m}/m\right)^{r\eta(\rho\pm\delta)}$ is uniformly integrable for any $r>1$, with choosing δ to satisfy $\rho\pm\delta \le 3/(r\eta)$. Therefore, it remains to prove that there exists some $r>1$ such that

$$\sup_{m>m_0} \mathbb{E}\left|g'\left(\ln\left(\frac{\hat{Z}_{2:2}-\hat{Z}_{1:2}}{\hat{Z}_{1:2}-\hat{Z}_{0:2}}\right)\right)\right|^{r\eta} < \infty. \tag{C.2}$$

Denote $\frac{b}{c} := \frac{Z_{2:2}-Z_{1:2}}{Z_{1:2}-Z_{0:2}}$ and $\frac{d}{e} := \frac{h(Y_{2:2};Y^*_{m-2:m})-h(Y_{1:2};Y^*_{m-2:m})}{h(Y_{1:2};Y^*_{m-2:m})-h(Y_{0:2};Y^*_{m-2:m})}$ Then $\frac{\hat{Z}_{2:2}-\hat{Z}_{1:2}}{\hat{Z}_{1:2}-\hat{Z}_{0:2}} = \frac{\hat{a}b+(1-\hat{a})d}{\hat{a}c+(1-\hat{a})e}$ is bounded by $\frac{b}{c}$ and $\frac{d}{e}$. By the assumed monotonicity of g' in Condition 5, we have

$$\left|g'\left(\ln\left(\frac{\hat{Z}_{2:2}-\hat{Z}_{1:2}}{\hat{Z}_{1:2}-\hat{Z}_{0:2}}\right)\right)\right|^{\eta r} \le \left|g'\left(\ln\left(\frac{Z_{2:2}-Z_{1:2}}{Z_{1:2}-Z_{0:2}}\right)\right)\right|^{\eta r} + \left|g'\left(\ln\left(\frac{h(Y_{2:2};Y^*_{m-2:m})-h(Y_{1:2};Y^*_{m-2:m})}{h(Y_{1:2};Y^*_{m-2:m})-h(Y_{0:2};Y^*_{m-2:m})}\right)\right)\right|^{\eta r}.$$

The latter inequality, together with Condition 5, implies the relation (C.2).

By combining the uniform integrability of $|\xi_m|^\eta \mathbb{I}_{D_m}$ and $|\xi_m|^\eta \mathbb{I}_{D^c_m}$, we conclude that ξ_m is uniformly integrable. It follows that

$$\begin{aligned}\lim_{m\to\infty}\mathbb{E}[\xi_m] &= \mathbb{E}\left[\lim_{m\to\infty}\xi_m\right] \\ &= \mathbb{E}\left[S_3^{-\rho}\right]\mathbb{E}\left[\sum_{j=1}^{3}\dot{K}_{(j)}\left(\{h_\gamma\left(Y_{3-j:2}\right)\}_{j=1}^{3}\right)H_{\gamma,\rho}\left(Y_{3-j,:2}\right)\right] = B_K(\rho,\gamma).\end{aligned}$$

Next, we turn to prove the second half of the condition: $\sum_{\ell=1}^{m} p_{n,m}(\ell)\Psi_{m,\ell} \to 0$ as $n \to \infty$, where $\Psi_{m,\ell} = \operatorname{Cov}\left(\xi_m^I, \xi_m^{I'}\right)$ for index sets such that $|I'| = |I| = m$ and $|I \cap I'| = \ell$. Note that all steps above in conjunction with Lemma C.3 can be adapted to prove uniform integrability of ξ_m^2, from which it follows that

$$\lim_{m\to\infty} \mathbb{E}[\xi_m^2] = \mathbb{E}\left[S_3^{-2\rho}\right] \mathbb{E}\left[\left(\sum_{j=1}^{3} \dot{K}_{(j)}\left(\{h_\gamma\left(Y_{3-j:2}\right)\}_{j=1}^{3}\right) H_{\gamma,\rho}\left(Y_{3-j:2}\right)\right)^2\right] =: \phi.$$

By classical theory on U-statistics [18, Chapter 5.2], we have

$$\sum_{\ell=1}^{m} p_{n,m}(\ell)\Psi_{m,\ell} \le \frac{m}{n}\Psi_{m,m},$$

where $\Psi_{m,m}$ is the variance of ξ_m. The second half of the condition follows immediately from $\mathbb{E}[\xi_m] \to B_K(\rho,\gamma)$ and $\mathbb{E}[\xi_m^2] \to \phi$ as $m \to \infty$. Together with Condition 4, we conclude that

$$\sqrt{k}\left(U_n^m - U_{n,Z}^m\right) \xrightarrow{p} \lambda B_K(\rho,\gamma), \qquad \text{as } n \to \infty.$$

□

Lemma C.3. *Recall* $\tilde{h}(Y_{i:2}) = h(Y_{i:2}; Y^*_{m-2:m}) - h_\gamma\left(Y_{i:2}\right)$ *and* $\hat{Z}_m = (1-\hat{a})\{h(Y_{3-j:2}; Y^*_{m-2:m})\}_{j=1}^3 + \hat{a}\{h_\gamma(Y_{3-j:2})\}_{j=1}^3$ *for some* $0 < \hat{a} < 1$*, where* $\hat{a}$ *is a random variable. Recall that* $\{\hat{Z}_{3-j:2}\}_{j=1}^3 := \hat{Z}_m$ *where* $\hat{Z}_{2:2} \ge \hat{Z}_{1:2} \ge \hat{Z}_{0:2}$ *and* $D_m = \{Y^*_{m-q+1:m} > t_0\}$*. For sufficiently large* $t_0 > 1$*, we have*

$$\sup_{m\ge 3} \sum_{j=1}^{2} \left| \frac{\tilde{h}(Y_{3-j:2}) - \tilde{h}(Y_{2-j:2})}{\left(\hat{Z}_{3-j:2} - \hat{Z}_{2-j:2}\right) A(Y^*_{m-2:m})} \right| \mathbb{I}_{D_m} \le B$$

for a random variable B *satisfying* $\mathbb{E}[|B|^r] < \infty$ *for any* $r > 0$.

Proof. For brevity, we only deal with the first term in the summation:

$$\left| \frac{\tilde{h}(Y_{2:2}) - \tilde{h}(Y_{1:2})}{(\hat{Z}_{2:2} - \hat{Z}_{1:2})\, A(Y^*_{m-2:m})} \right|.$$

The second term can be handled using the same inequalities even though $Y_{0:2} = 1$.

Recall the function $A(t)$ from Condition 3 as

$$A(t) = \frac{t\, U''(t)}{U'(t)} - \gamma + 1.$$

Following the proof of Theorem 2.3.12 in [4], we can write

$$\frac{\tilde{h}(Y_{2:2}) - \tilde{h}(Y_{1:2})}{A(Y^*_{m-2:m})} = \int_{Y_{1:2}}^{Y_{2:2}} s^{\gamma-1} \int_1^s \frac{A(Y^*_{m-2:m}u)}{A(Y^*_{m-2:m})} \frac{U'(Y^*_{m-2:m}u)}{U'(Y^*_{m-2:m})} u^{-\gamma}\, \mathrm{d}u\, \mathrm{d}s.$$

By regular variation of the absolute value of A and U' with indices ρ and $\gamma - 1$ respectively and using Potter's bounds in Lemma C.1, we get that for $\delta_2 > 0$ that there exists a $t_0(\delta_2)$ such that if $Y^*_{m-2:m} > t_0(\delta_2)$,

$$\begin{aligned}
&\int_{Y_{1:2}}^{Y_{2:2}} s^{\gamma-1} \int_1^s \frac{A(Y^*_{m-2:m}u)}{A(Y^*_{m-2:m})} \frac{U'(Y^*_{m-2:m}u)}{U'(Y^*_{m-2:m})} u^{-\gamma} \, \mathrm{d}u \, \mathrm{d}s \\
&\le \int_{Y_{1:2}}^{Y_{2:2}} s^{\gamma-1} \int_1^s (1+\delta_2) u^{\rho-1+\delta_2} \, \mathrm{d}u \, \mathrm{d}s \\
&= \frac{1+\delta_2}{\rho+\delta_2} \left(h_{\gamma+\rho+\delta_2}(Y_{2:2}) - h_{\gamma+\rho+\delta_2}(Y_{1:2}) \right) - \frac{1+\delta_2}{\rho+\delta_2} \left(h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2}) \right),
\end{aligned}$$

provided $\rho + \delta_2 \neq 0$, and similarly,

$$\begin{aligned}
&\int_{Y_{1:2}}^{Y_{2:2}} s^{\gamma-1} \int_1^s \frac{A(Y^*_{m-2:m}u)}{A(Y^*_{m-2:m})} \frac{U'(Y^*_{m-2:m}u)}{U'(Y^*_{m-2:m})} u^{-\gamma} \, \mathrm{d}u \, \mathrm{d}s \\
&\ge \frac{1-\delta_2}{\rho-\delta_2} \left(h_{\gamma+\rho-\delta_2}(Y_{2:2}) - h_{\gamma+\rho-\delta_2}(Y_{1:2}) \right) - \frac{1-\delta_2}{\rho-\delta_2} \left(h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2}) \right).
\end{aligned}$$

Accordingly, it suffices to prove that, for any $r > 0$ and $\gamma \in \mathbb{R}$, the term

$$\left| \frac{h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2})}{\hat{Z}_{2:2} - \hat{Z}_{1:2}} \right|^r$$

can be bounded by an integrable function not depending on m.

Note that

$$\begin{aligned}
\hat{Z}_{2:2} - \hat{Z}_{1:2} = (1-\hat{a}) \big(h(Y_{2:2}; Y^*_{m-2:m}) - h(Y_{1:2}; Y^*_{m-2:m}) \big) \\
+ \hat{a} \left(h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2}) \right).
\end{aligned}$$

We find, if $Y^*_{m-2:m} > t_0(\gamma, \delta_1)$, using Potter's bounds in Lemma C.1,

$$\begin{aligned}
h(Y_{2:2}; Y^*_{m-2:m}) - h(Y_{1:2}; Y^*_{m-2:m}) &= \int_{Y_{1:2}}^{Y_{2:2}} \frac{U'(tY^*_{m-2:m})}{U'(Y^*_{m-2:m})} \, \mathrm{d}t \\
&\ge (1-\delta_1) \int_{Y_{1:2}}^{Y_{2:2}} t^{\gamma-1-\delta_1} \, \mathrm{d}t = (1-\delta_1) \left(h_{\gamma-\delta_1}(Y_{2:2}) - h_{\gamma-\delta_1}(Y_{1:2}) \right),
\end{aligned}$$

provided $\gamma - \delta_1 \neq 0$. Define $\tilde{Y} = Y_{2:2}/Y_{1:2}$. Recalling that $0 < \hat{a} < 1$, we find

$$\begin{aligned}
&\hat{Z}_{2:2} - \hat{Z}_{1:2} \\
&\ge (1-\hat{a})(1-\delta_1) \left(h_{\gamma-\delta_1}(Y_{2:2}) - h_{\gamma-\delta_1}(Y_{1:2}) \right) + \hat{a} \left(h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2}) \right) \\
&\ge (1-\hat{a})(1-\delta_1) \left(h_{\gamma-\delta_1}(Y_{2:2}) - h_{\gamma-\delta_1}(Y_{1:2}) \right) + \hat{a}(1-\delta_1) \left(h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2}) \right) \\
&\ge (1-\delta_1) Y_{2:2}^{-\delta_1} \left(h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2}) \right),
\end{aligned}$$

where the last inequality follows from the fact that for $0 < x < y < \infty$, for $\gamma \in \mathbb{R}$ and for $0 \le \delta < \infty$, we have

$$y^\delta \left(h_{\gamma-\delta}(y) - h_{\gamma-\delta}(x)\right) = y^\delta \int_x^y t^{\gamma-\delta-1}\,\mathrm{d}t = \int_x^y t^{\gamma-1}(y/t)^\delta\,\mathrm{d}t$$
$$\ge \int_x^y t^{\gamma-1}\,\mathrm{d}t = h_\gamma(y) - h_\gamma(x).$$

We conclude with the bound

$$\left|\frac{h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2})}{\hat{Z}_{2:2} - \hat{Z}_{1:2}}\right| \le \frac{h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2})}{(1-\delta_1)Y_{2:2}^{-\delta_1}\left(h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2})\right)} \le \frac{Y_{2:2}^{\delta_1}}{(1-\delta_1)},$$

from which the lemma follows by choosing $\delta_1 < (2r)^{-1}$. ◻

Appendix D: Proofs for Section 4

Proof of Lemma 4.1. The formulas for $\mathbb{E}[\ln Z_1]$, $\mathbb{E}[\ln Z_{1:2}]$ and $\mathbb{E}[\ln X_{2:2}]$ follow directly from the moments of exponentiated GP random variables and their order statistics, see [13, Section 2.3].

Let Y_1, Y_2 be two independent standard Pareto random variables and let $Y_{1:2} \le Y_{2:2}$ denote their order statistics as. Recall that we can represent the $\mathrm{GP}(\gamma)$ random variables Z_1, Z_2 and their corresponding order statistics $Z_{1:2} \le Z_{2:2}$ as $Z_i = h_\gamma(Y_i)$, $Z_{1:2} = h_\gamma(Y_{1:2})$ and $Z_{2:2} = h_\gamma(Y_{1:2})$. But then

$$Z_{2:2} - Z_{1:2} = h_\gamma(Y_{2:2}) - h_\gamma(Y_{1:2}) = Y_{1:2}^\gamma h_\gamma(Y_{2:2}/Y_{1:2}).$$

Since $Y_{2:2}/Y_{1:2}$ is equal in distribution to Y_1, we find

$$u_1(\gamma) = \mathbb{E}[\ln Y_{1:2}^\gamma] + \mathbb{E}[\ln Z_1] - \mathbb{E}[\ln Z_{1:2}]$$
$$= \begin{cases} \gamma/2 - \psi(1/\gamma) + \psi(2/\gamma) & \text{if } \gamma > 0 \\ \gamma/2 - \psi(1-1/\gamma) + \psi(1-2/\gamma) & \text{if } \gamma < 0 \\ \ln(2) & \text{if } \gamma = 0. \end{cases}$$

More straightforwardly, we have

$$u_2(\gamma) = \mathbb{E}[\ln Z_{2:2}] - \mathbb{E}[\ln Z_{1:2}]$$
$$= \begin{cases} 2\left(\psi(2/\gamma) - \psi(1/\gamma)\right) & \text{if } \gamma > 0 \\ 2\left(\psi(1-2/\gamma) - \psi(1-1/\gamma)\right) & \text{if } \gamma < 0 \\ 2\ln(2) & \text{if } \gamma = 0. \end{cases}$$

It follows that $2u_1(\gamma) - u_2(\gamma) = \gamma$, as required. ◻

Proof of Proposition 4.3. One can check that the kernel K_m^* is differentiable, has continuous first-order partial derivatives and that the auxiliary function g

satisfies $|g'(x)| = |2 - e^x/(e^x+1)| \le 2$. Therefore, Corollary 3.2 is applicable if K_m^* satisfies Conditions 1 and 6.

Condition 1 can be verified as follows: for any $p > 0$, with the same notation as in the proof of Lemma 4.1,

$$\begin{aligned}
\mathbb{E}\left[|K_m^*(Z_1,\ldots,Z_m)|^p\right] &= \mathbb{E}\left[\left|2\ln\left(\frac{Z_{2:2}-Z_{1:2}}{Z_{1:2}}\right) - \ln\left(\frac{Z_{2:2}}{Z_{1:2}}\right)\right|^p\right] \\
&= \mathbb{E}\left[\left|2\ln\left[Y_{1:2}^{\gamma}\right] + 2\ln\left[h_\gamma(Y_{2:2}/Y_{1:2})\right] - \ln\left[h_\gamma(Y_{1:2})\right] - \ln\left[h_\gamma(Y_{2:2})\right]\right|^p\right] < \infty,
\end{aligned}$$

because exponential random variables and exponentiated GP random variables (and its order statistics) have finite moments of all orders [13, Corollary 3].

For Condition 6, recall that K_m^* is a linear combination of the kernels $K_{m,1}$ and $K_{m,2}$ in (4.1) which themselves consist of linear combinations of the logarithms of the spacings $X_{m:m} - X_{m-1:m}$, $X_{m:m} - X_{m-2:m}$ and $X_{m-1:m} - X_{m-2:m}$. Note that

$$\begin{aligned}
&-\ln(X_{m:m} - X_{m-2:m})\,\mathbb{1}_{\{X_{m:m}-X_{m-2:m}<1\}} \\
&\qquad\qquad \le -\ln(X_{m:m} - X_{m-1:m})\,\mathbb{1}_{\{X_{m:m}-X_{m-1:m}<1\}}
\end{aligned}$$

and for $0 \le j < i \le 2$,

$$\begin{aligned}
&\ln(X_{m-j:m} - X_{m-i:m})\,\mathbb{1}_{\{X_{m-j:m}-X_{m-i:m}>1\}} \\
&\qquad\qquad \le \ln(X_{m:m} - X_{m-2:m})\,\mathbb{1}_{\{X_{m:m}-X_{m-2:m}>1\}}\,.
\end{aligned}$$

Consequently, to verify Condition 6 we only have to show that

$$\mathbb{E}\left[(-\ln(X_{m-j:m} - X_{m-j-1:m}))^p\,\mathbb{1}_{\{X_{m-j:m}-X_{m-j-1:m}<1\}}\right] = \mathrm{O}(m^a) \tag{D.1}$$

for $j \in \{0,1\}$, and

$$\mathbb{E}\left[(\ln(X_{m:m} - X_{m-2:m}))^p\,\mathbb{1}_{\{X_{m:m}-X_{m-2:m}>1\}}\right] = \mathrm{O}(m^a) \tag{D.2}$$

for some $p > 2$ and $a > 0$.

First, for the expectation in (D.1), note that the distribution function $\tilde{F}_j$ of the spacing $X_{m-j:m} - X_{m-j-1:m}$ for $j \in \{0,1\}$ is given by

$$\tilde{F}_j(r) = 1 - \int_{-\infty}^{+\infty} (m-j-1)\binom{m}{m-j-1} f(x) F^{m-j-2}(x)\,(1 - F(x+r))^{j+1}\,\mathrm{d}x.$$

Moreover, note that $\left|(1-F(x))^2 - (1-F(a))^2\right| \le 4\,|F(x) - F(a)|$ for all $x, a \in$

$\mathcal{E}$ and $\tilde{F}_j(0) = 0$. We get, for any $0 < x_0 < 1$ and $r_0 = (-\ln x_0)^p$, that

$$\begin{aligned}
&\mathbb{E}\left[(-\ln(X_{m-j:m} - X_{m-j-1:m}))^p \, \mathbb{1}_{\{X_{m-j:m}-X_{m-j-1:m}<x_0\}}\right] \\
&= \int_{r_0}^{\infty} \mathbb{P}(X_{m-j:m} - X_{m-j-1:m} < e^{-r^{1/p}}) \, \mathrm{d}r \\
&= \int_{r_0}^{\infty} \left(\tilde{F}_j(e^{-r^{1/p}}) - \tilde{F}_j\,(0)\right) \mathrm{d}r \\
&\leq (m-j-1)\binom{m}{m-j-1} \\
&\quad \cdot \int_{r_0}^{\infty} \int_{-\infty}^{+\infty} f(x) F^{m-j-2}(x) \left| \left(1 - F(x + e^{-r^{1/p}})\right)^{j+1} - (1 - F(x))^{j+1} \right| \mathrm{d}x \, \mathrm{d}r \\
&\leq c\binom{m}{m-j-1} \int_{r_0}^{\infty} r^{-(1+\epsilon)} \mathrm{d}r,
\end{aligned}$$

for some constant $c > 0$. Here the last step follows from condition (4.5). With straightforward calculation, we get that as $m \to \infty$,

$$\mathbb{E}\left[(-\ln(X_{m-j:m} - X_{m-j-1:m}))^p \, \mathbb{1}_{\{X_{m-j:m}-X_{m-j-1:m}<1\}}\right] = \mathrm{O}(m^2).$$

Secondly, for the expectation in (D.2), consider the following five events:

$$\begin{aligned}
A &:= \{X_{m:m} - X_{m-2:m} > 1\}, \\
A_1 &:= \{X_{m:m} > 1, X_{m-2:m} < -1\}, \\
A_2 &:= \{X_{m:m} \leq 1, X_{m-2:m} < -1\}, \\
A_3 &:= \{X_{m:m} > 1, X_{m-2:m} \geq -1\}, \\
A_4 &:= \{X_{m:m} \leq 1, X_{m-2:m} \geq -1\}.
\end{aligned}$$

The following inequalities can be verified in a straightforward way:

$$\begin{aligned}
&(\ln(X_{m:m} - X_{m-2:m}))^p \, \mathbb{1}_{A\cap A_1} \leq 2^p \left[(\ln(X_{m:m}))^p + (\ln(-X_{m-2:m}))^p + 1\right] \mathbb{1}_{A\cap A_1}, \\
&(\ln(X_{m:m} - X_{m-2:m}))^p \, \mathbb{1}_{A\cap A_2} \leq 2^p \left[(\ln(-X_{m-2:m}))^p + 1\right] \mathbb{1}_{A\cap A_2}, \\
&(\ln(X_{m:m} - X_{m-2:m}))^p \, \mathbb{1}_{A\cap A_3} \leq 2^p \left[(\ln(X_{m:m}))^p + 1\right] \mathbb{1}_{A\cap A_3}, \\
&(\ln(X_{m:m} - X_{m-2:m}))^p \, \mathbb{1}_{A\cap A_4} \leq \mathbb{1}_{A\cap A_4}.
\end{aligned}$$

Therefore, we only need to show that the following two expectations

$$\mathbb{E}\left[(\ln(X_{m:m}))^p \, \mathbb{1}_{\{X_{m:m}>1\}}\right] \quad \text{and} \quad \mathbb{E}\left[(\ln(-X_{m-2:m}))^p \, \mathbb{1}_{\{X_{m-2:m}<-1\}}\right],$$

are of the order $\mathrm{O}(m^a)$ for some constant $a > 0$.

We start from the second expectation concerning $X_{m-2:m}$. There exists $x_0 > 1$ such that for all $m > 3$, the variable $(-X_{m-2:m})\,\mathbb{1}_{\{X_{m-2:m}<-x_0\}}$ is stochastically dominated by $(-X)\,\mathbb{1}_{\{X<-x_0\}}$, where X has distribution F. Therefore,

$$\begin{aligned}
&\mathbb{E}\left[(\ln(-X_{m-2:m}))^p \, \mathbb{1}_{\{X_{m-2:m}<-1\}}\right] \\
&\qquad\qquad \leq \mathbb{E}\left[(\ln(-X))^p \, \mathbb{1}_{\{X<-x_0\}}\right] + (\ln(x_0))^p = \mathrm{O}(1).
\end{aligned}$$

Finally, for the first expectation concerning $X_{m:m}$, choose $0 < \theta < 1/\gamma$. Note that there exists $x_0 = x_0(p) > 1$ such that $(\ln(x))^p$ is a concave function in $x > x_0$. In view of Jensen's inequality, we have

$$\mathbb{E}\left[(\ln(X_{m:m}))^p \mathbb{1}_{\{X_{m:m}>x_0\}}\right] \leq \left(\frac{1}{\theta}\ln\left(\mathbb{E}\left[X_{m:m}^{\theta}\mathbb{1}_{\{X_{m:m}>x_0\}}\right]\right)\right)^p.$$

For $\theta < 1/\gamma$, Theorem 5.3.2 in [4] yields that as $m \to \infty$,

$$\mathbb{E}\left[X_{m:m}^{\theta}\mathbb{1}_{\{X_{m:m}>1\}}\right] = (U(m))^{\theta} = \mathrm{o}(m),$$

which is sufficient to bound the first expectation by $\mathrm{O}(m^a)$ with any $a > 0$. Combining the two expectations completes the proof of (D.2) and thus that of Condition 6.

Finally, we show that the condition

$$\sup_{x\in\mathcal{E}}\left|F(x) - F(x + e^{-r^{1/p}})\right| = \mathrm{O}(r^{-(1+\epsilon)}), \qquad \text{as } r \to \infty,$$

is satisfied if the density $f(x)$ is bounded. By the mean value theorem there exists a point $c := c(x, r)$ between x and $x + e^{-r^{1/p}}$ such that

$$F(x + e^{-r^{1/p}}) - F(x) = f(c)e^{-r^{-1/p}}.$$

Taking the supremum with respect to x on both sides of the previous equality then immediately implies the result by boundedness of f. $\square$

Proof of Lemma 4.4. Put $S_{ij} = \ln(X_{n-i+1:n} - X_{n-j+1:n})$ for integer $1 \leq i < j \leq n$. For $I \subset [n]$ with $|I| = m$, both $K_{m,1}(X_I)$ and $K_{m,2}(X_I)$ are linear combinations of such log-spacings S_{ij}. For a given pair (i, j) of indices, the number of such subsets I in which S_{ij} appears can be counted:

- There are $\binom{n-j}{m-2}$ sets I such that $X_{n-i+1:n}$ and $X_{n-j+1:n}$ are respectively the largest and 2nd largest order statistics of X_I.
- There are $(i-1)\binom{n-j}{m-3}$ sets I such that $X_{n-i+1:n}$ and $X_{n-j+1:n}$ are respectively the 2nd and 3rd largest order statistics of X_I.
- There are $(j-i-1)\binom{n-j}{m-3}$ sets I such that $X_{n-i+1:n}$ and $X_{n-j+1:n}$ are respectively the largest and 3rd largest order statistics of X_I.

Since $n - j + 1$ cannot be lower than $m - 2$ (because there need to be at least $m - 3$ observations lower than $X_{n-j+1:n}$), we find

$$U_{n,*}^m = 2U_{n,1}^m - U_{n,2}^m$$

with

$$U_{n,1}^m = \binom{n}{m}^{-1}\sum_{j=2}^{n-m+3}\sum_{i=1}^{j-1}\left[\binom{n-j}{m-2} - (i-1)\binom{n-j}{m-3}\right]S_{ij},$$

$$U_{n,2}^m = \binom{n}{m}^{-1}\sum_{j=2}^{n-m+3}\sum_{i=1}^{j-1}\left[(j-i-1)\binom{n-j}{m-3} - (i-1)\binom{n-j}{m-3}\right]S_{ij}.$$

It follows that

$$U_{n,*}^m = \binom{n}{m}^{-1} \sum_{j=2}^{n-m+3} \left[2\binom{n-j}{m-2} - j\binom{n-j}{m-3} \right] \sum_{i=1}^{j-1} S_{ij},$$

which can be further simplified to the stated formula. □

References

[1] Bücher, A. and Segers, J. (2018). Inference for heavy tailed stationary time series based on sliding blocks. *Electron. J. Statist.* **12** 1098–1125.

[2] Chvátal, V. (1979). The tail of the hypergeometric distribution. *Discrete Mathematics* **25** 285–287.

[3] Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)* **52** 393–425.

[4] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.

[5] Dombry, C. and Ferreira, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli* **25** 1690–1723.

[6] Drees, H., Ferreira, A. and De Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *Annals of Applied Probability* 1179–1201.

[7] Ferreira, A. and De Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *The Annals of statistics* **43** 276–298.

[8] Grimshaw, S. D. (1993). Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics* **35** 185–191.

[9] Halmos, P. R. (1946). The theory of unbiased estimation. *The Annals of Mathematical Statistics* **17** 34–43.

[10] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19** 293–325.

[11] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58** 13–30.

[12] Lee, A. J. (1990). *U-statistics. Theory and practice*. Marcel Dekker, Inc., New York.

[13] Lee, S. and Kim, J. H. (2019). Exponentiated generalized Pareto distribution: Properties and applications towards extreme value theory. *Communications in Statistics – Theory and Methods* **48** 2014–2038.

[14] Oorschot, J. and Zhou, C. (2020). All Block Maxima method for estimating the extreme value index. *arXiv preprint arXiv:2010.15950*.

[15] Robert, C. Y., Segers, J. and Ferro, C. A. T. (2009). A sliding blocks estimator for the extremal index. *Electron. J. Stat.* **3** 993–1020. MR2540849

[16] Segers, J. (2001). Extreme of a random sample: limit theorems and statistical applications, PhD thesis, Katholieke Universiteit Leuven.

[17] Segers, J. and Teugels, J. (2000). Testing the Gumbel hypothesis by Galton's ratio. *Extremes* **3** 291–303.

[18] Serfling, R. J. (2009). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

[19] van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

[20] Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index. *J. Multivar. Anal.* **100** 794–815.