Detecting tail behavior: mean excess plots with confidence bounds

Bikramjit Das¹ · Souvik Ghosh²

Received: 11 November 2014 / Revised: 12 November 2015 / Accepted: 21 December 2015 / Published online: 7 January 2016
© Springer Science+Business Media New York 2016

Abstract In many practical situations exploratory plots are helpful in understanding tail behavior of sample data. The Mean Excess plot is one of the exploratory tools often used in practice to understand the right tail behavior of a data set. It is known that if the underlying distribution of a data sample is in the maximum domain of attraction of a Fréchet, a Gumbel or a Weibull distributions then the ME plot of the data approaches a straight line in an appropriate sense, with positive, zero or negative slope respectively. In this paper we construct confidence intervals around the ME plots which assist us in ascertaining which particular maximum domain of attraction the data set comes from. We recall weak limit results for the Fréchet domain of attraction, already obtained in Das and Ghosh (Bernoulli 19, 308–342 2013) and derive weak limits for the Gumbel and Weibull domains in order to construct confidence bounds. We demonstrate our methodology by applying them to simulated and real data sets.

Keywords Extreme values · Regular variation · Random set · ME plot · Asymptotic theory · Confidence bounds

Bikramjit Das gratefully acknowledges support from SRG ESD 047 and MIT-SUTD IDC grant IDG31300110. The authors are grateful to the anonymous referees for their valuable comments which have helped improve the article.

✉ Bikramjit Das
bikram@sutd.edu.sg

Souvik Ghosh
sghosh@linkedin.com

¹ Pillar of Engineering Systems and Design, Singapore University of Technology and Design, 8 Somapah Drive, Singapore 487372, Singapore

² LinkedIn Corporation, 2029 Stierlin Court, Mountain View, CA 94043, USA
AMS 2000 Subject Classifications  Primary 62G32, 62-09, 60G70 · Secondary 62G10, 62G15, 60F05

1 Introduction

This article concerns the use of Mean Excess plot or ME plot, a popular exploratory tool used to understand the tail behavior of a univariate data set. Given a sample of data points, one of the first things a sensible data analyst does is to compute a summary statistics. Such a summary statistic might involve calculating measures of central tendencies (mean, median, mode) and those of dispersion (standard deviation, range, etc), plotting a histogram, an empirical cumulative distribution function and so on and so forth. A more curious analyst would ask the question, does it even make sense to calculate the sample mean or standard deviation; would they represent their counterparts in the original population? What if the probability distribution of the population from which the data is sampled does not even have a first or second moment. This is a question that would or perhaps should particularly come to the mind of analysts modeling risk or other extreme events. In a world where data is being used to make serious economic, financial or environmental policy decisions, understanding extreme risks, which relate to the tail behavior of data sets have become increasingly important. This can be easily observed in the world of finance and insurance (Das et al. 2013; Donnelly and Embrechts 2010; McNeil et al. 2005), telecommunications (Maulik et al. 2002), environmental statistics (Davison and Smith 1990) and many more areas.

The ME plot is a graphical tool that is widely used to understand the tail behavior of a sample; see Embrechts et al. (1997); Davison and Smith (1990); Drees (2012); Neves and Fraga Alves (2008). The ME plot, if the mean exists, assists in distinguishing light-tailed data sets from heavy-tailed ones. The inference is based on a visual examination of the slope of a fitted line through the ME plot (to be described in the next section) being zero, less than zero or greater than zero. Clearly, a confidence set around the fitted line would make inference in these cases more meaningful; hence this is the aim of the paper.

1.1 The ME plot

The ME plot, as described in the introduction, is a popular tool in extreme value analysis. It is a simple graphical test to check whether data conform to a generalised Pareto distribution (GPD). The class of GPD arise naturally in extreme value analysis as limit distributions while using the peaks-over-threshold (POT) method (Beirlant et al. 2004; Davison and Smith 1990). The cumulative distribution function of a GPD is:

\[
G_{\xi, \beta}(x) = \begin{cases} 
1 - (1 + \frac{\xi x}{\beta})^{-1/\xi} & \text{if } \xi \neq 0, \\
1 - \exp(-x/\beta) & \text{if } \xi = 0,
\end{cases} \tag{1.1}
\]
where $\beta > 0$, and $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ if $\xi < 0$. Parameters $\xi$ and $\beta$ are referred to as the *shape* and the *scale* parameter respectively. In extreme value analysis, we are interested in the shape parameter $\xi$ which tells us whether the data is heavy-tailed ($\xi > 0$) or light-tailed ($\xi \leq 0$) or even more specifically if the underlying distribution has finite right end-point ($\xi < 0$). The case $\xi > 0$ and $\beta = 1$ corresponds to the classical Pareto law with tail exponent $1/\xi$.

The ME plot is an empirical graphical plot of the *ME function* of a random variable $X \sim F$ which is defined as:

$$M(u) := \mathbb{E}[X - u|X > u],$$

provided $\mathbb{E}X_+ < \infty$. The ME function is also known as the *mean residual life function* for non-negative random variables and is extensively used in reliability theory and survival analysis for data modelling since $M(u)$ completely determines $F$ if $\mathbb{E}(X) < \infty$ (Hall and Wellner 1981). Suppose we have an iid sample $X_1, \ldots, X_n \sim F$. A natural estimate of $M(u)$ is the empirical ME function $\hat{M}(u)$ defined as

$$\hat{M}(u) = \frac{\sum_{i=1}^{n}(X_i - u)I[X_i > u]}{\sum_{i=1}^{n}I[X_i > u]}, \quad u \geq 0. \quad (1.3)$$

Denoting $X(1) \geq X(2) \geq \ldots \geq X(n)$ to be the order statistics from a sample $X_1, \ldots, X_n$, the ME plot is a plot of the points $\mathcal{ME}_n := \{(X(k), \hat{M}(X(k))): 1 < k \leq n\}$.

We study the asymptotic properties of $\mathcal{ME}_n$ for different classes of distributions $F$. It is well-known that for a random variable $X \sim G_{\xi, \beta}$, we have $E(X) < \infty$ if and only if $\xi < 1$ and in this case, the ME function of $X$ is linear in $u$:

$$M(u) = \frac{\beta}{1 - \xi} + \frac{\xi}{1 - \xi}u, \quad (1.4)$$

where $0 \leq u < \infty$ if $0 \leq \xi < 1$ and $0 \leq u \leq -\beta/\xi$ if $\xi < 0$.

Interestingly, the linearity of the ME function also characterises the *GPD* class (McNeil et al. 2005; Embrechts et al. 1997). From the discussions in Ghosh and Resnick (2010) we know that the empirical ME plot $\mathcal{ME}_n$ above a high order statistic $X(k)$, when appropriately normalised converge in probability to a straight line if $F$ is in the maximum domain of attraction of a generalized extreme value distribution with finite mean (Gumbel, Weibull or Fréchet distribution). Distributional limits for $\mathcal{ME}_n$ in a space of closed sets and confidence intervals around $\mathcal{ME}_n$ can also be computed in many cases and such findings have been discussed in case the underlying data is heavy-tailed (Fréchet domain of attraction) in Das and Ghosh (2013).

In Section 1.2 we collect notations and ideas to be used throughout the paper. See Das and Ghosh (2013) for further elaboration on the concepts of convergence of closed sets. In the main part of the paper we start by consolidating a few known distributional properties of ME plot, especially in the heavy-tailed case (Das and Ghosh 2013); this is covered in Section 2.1. The rest of Section 2 deals with limit results for ME plots in the case where the underlying distribution is either in the Gumbel
maximum domain of attraction or in a Weibull maximum domain of attraction. The limit theorems proved in Section 2 are used to create confidence bounds around the ME plots in Section 3. In Section 4 we use the tools developed in Sections 2 and 3 to detect and test tail behavior for simulated and real data sets.

1.2 Miscellany

First we recall the idea of maximum domain of attraction of an extreme value distribution. The class of extreme value distributions is parametrized by a shape parameter \( \xi \in \mathbb{R} \) and we define the distribution function \( G_\xi \) to be

\[
G_\xi (x) = \exp(- (1 + \xi x)^{-1/\xi}), \quad 1 + \xi x > 0,
\]

for all real \( \xi \) and for \( \xi = 0 \), the right hand side is interpreted as \( \exp(-e^{-x}) \).

**Definition 1.1** A distribution function \( F \) (or the underlying random variable \( X \sim F \)) is in the maximum domain of attraction of an extreme value distribution \( G_\xi \) if there exists sequences \( c_n > 0 \) and \( d_n \in \mathbb{R} \) such that

\[
F^n(c_n x + d_n) \rightarrow G_\xi (x) \quad \text{for all } x \in \mathbb{R}.
\]

The distributions for the cases \( \xi > 0, \xi = 0 \) and \( \xi < 0 \) are respectively called the Fréchet distribution, the Gumbel distribution and the Weibull distribution. As mentioned in the introduction, if \( F \in D(G_\xi) \) for some extreme value distribution with \( \xi < 1 \), implying that \( F \) has finite mean, then the ME function of \( F \) is linear with an appropriate slope determined by the parameter \( \xi \); see Ghosh and Resnick (2010).

Throughout this paper we will take \( k := k_n \) to be a sequence increasing to infinity such that \( n/k_n \rightarrow \infty \) or \( k_n/n \rightarrow 0 \). For a distribution function \( F(x) \) we write \( \bar{F}(x) := 1 - F(x) \) and the tail quantile function

\[
b(u) := F^\rightarrow((1 - \frac{1}{u}) := \inf \left\{ s : F(s) \geq 1 - \frac{1}{u} \right\} = \left( \frac{1}{1 - F} \right)^\leftarrow(u).
\]

A function \( U : (0, \infty) \rightarrow \mathbb{R}_+ \) is regularly varying with index \( \rho \in \mathbb{R} \), written \( U \in RV_{\rho} \), if

\[
\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\rho, \quad x > 0.
\]

If \( X \sim F \) we will often have the right-hand tail of \( F \) to be regularly varying, that is, \( \tilde{F} \in RV_{-\alpha} \) for \( \alpha \geq 0 \), and by abuse of notation we might say \( X \in RV_{-\alpha} \). Regular variation is discussed in several books such as Resnick (2007, 2008; Seneta (1976); Geluk and de Haan (1987); de Haan (1970); de Haan and Ferreira (2006); Bingham et al. (1987).

We use \( \mathbb{M}_+(0, \infty) \) to denote the space of nonnegative Radon measures \( \mu \) on \( (0, \infty) \) metrized by the vague metric. Point measures are written as a function of their points \( \{x_i, i = 1, \ldots, n\} \) by \( \sum_{i=1}^n \delta_{x_i} \). See, for example, (Resnick 2008, Chapter 3).
We will use the following notations to denote different classes of functions: For $0 \leq a < b \leq \infty$,

1. $C[a, b)$: Continuous functions on $[a, b)$.
2. $D[a, b)$: Right-continuous functions with finite left limits defined on $[a, b)$.
3. $D_l[a, b)$: Left-continuous functions with finite right limits defined on $[a, b)$.

It is known that $D[0, 1]$ is complete and separable under a metric $d_0(\cdot)$ which is equivalent to the Skorohod metric $d_S(\cdot)$ (Billingsley 1968, p.128), but not under the uniform metric $\| \cdot \|$. As we will see, the limit processes that appear in our analysis below are always continuous. We can check that if $x$ is continuous (in fact uniformly continuous) in $[0, 1]$, for $x_n \in D[0, 1]$, $\|x_n - x\| \to 0$ is equivalent to $d_S(x_n, x) \to 0$ and hence equivalent to $d_0(x_n, x) \to 0$ as $n \to \infty$ (Billingsley 1968, p.124). So we use convergence in uniform metric, for our convenience henceforth. For spaces of the form $D[a, b)$ or $D_l[a, b)$ we will consider the topology of local uniform convergence. In some cases we will also consider product spaces of functions and then the topology will be the product topology. For example, $D_l^2[1, \infty)$ will denote the class of 2-dimensional functions on $[1, \infty)$ which are left-continuous with right limit. The classes of functions defined on the sets $[a, b)$ or $(a, b]$ will have the obvious notation. For further details on notions of convergence and topology in the context of convergences of plots see Das and Ghosh (2013).

2 Limit results for ME plots

In this section we find distributional limit for ME plots when they exist. We continue the study of ME plots from Ghosh and Resnick (2010) and Das and Ghosh (2013) and give a complete picture of limit results for ME plots. We cite some of the results from the afore-mentioned papers for completeness.

The basic assumption is that we have an iid sample of data points from some unknown distribution $F$ which belongs to the maximum domain of attraction of one of the three extreme value distributions. The assumption of independence in the sample can be relaxed a bit under certain conditions which we do not explore here.

Suppose $X_1, \ldots, X_n$ is an i.i.d. sample from a distribution $F$. We will work under this assumption for the entire section. The properties of the empirical ME function $\hat{M}(u)$ as an estimator of $M(u)$ has been studied by Yang (1978). It was shown there that $\hat{M}(u)$ is uniformly strongly consistent for $M(u)$: for any $0 < b < \infty$

$$P \left[ \lim_{n \to \infty} \sup_{0 \leq u \leq b} \left| \hat{M}(u) - M(u) \right| = 0 \right] = 1.$$ 

A weak (distributional) limit for $\hat{M}(u)$ was also shown in Yang (1978): for any $0 < b < 1$

$$\sqrt{n} \left( \hat{M} \left( F^{(u)}(t) \right) - M \left( F^{(u)}(t) \right) \right) \Rightarrow U(t),$$
where $U(t)$ is a Gaussian process on $[0, b]$ with covariance function

$$\Gamma(s, t) = \frac{(1-t)\sigma^2(t) - t\theta^2(t)}{(1-s)(1-t)^2}$$

for all $0 \leq s \leq t \leq b$

with

$$\sigma^2(t) = \text{var} \left( X I_{[t<F(X)\leq 1]} \right) \quad \text{and} \quad \theta(t) = E \left( X I_{[t<F(X)\leq 1]} \right).$$

Using Lemma 2.4 in Das and Ghosh (2013) it is easy to see that the ME plots also exhibit the same features. Our interest in ME plots is for detecting right tail behavior of data samples (an equivalent case can be easily made for left tail behavior). Hence the linearity we seek in the ME plot will be for high thresholds. Necessarily, the ME plots we will discuss in the various cases will be transformations of the ME plot above an appropriate quantile, i.e., $\{(X(i), \hat{M}(X(i)) : 1 < i \leq k \}$ for $k := k(n) < n$ where $\hat{M}$ is as defined in (1.3).

2.1 ME plot in the Fréchet case

First we look at the case where the underlying distribution $F$ is heavy-tailed, in the sense that $F \in D(G_\xi)$ with $\xi > 0$ or in other words, $\hat{F} \in RV_{-1/\xi}$. We define the ME plot as:

$$\mathcal{M}_n := \frac{1}{X(k)} \left\{ \left( X(i), \hat{M}(X(i)) \right) : i = 2, \ldots, k \right\}. \quad (2.1)$$

From (Ghosh and Resnick 2010, Theorem 3.2), we know that for $0 < \xi < 1$, as $n, k, n/k \to \infty$,

$$\mathcal{M}_n \xrightarrow{P} \mathcal{M} := \left\{ \left( t, \frac{\xi}{1-\xi} t \right) : t \geq 1 \right\}.$$

The distributional behavior of $\mathcal{M}_n$ depends on whether $F$ has finite second moment or not and has been discussed under certain regularity conditions in Das and Ghosh (2013). We note them down below but refer to Theorems 4.3 and 4.6 in Das and Ghosh (2013) for full details.

Case 1 ($0 < \xi < 1/2$): For any $0 < \epsilon < 1$ as $n, k, n/k \to \infty$,

$$\mathcal{M}_n^\epsilon := \left\{ \left( \left( \frac{i}{k} \right)^{-\xi}, \frac{\xi}{1-\xi} \left( \frac{i}{k} \right)^{-\xi} \right) : i = \lfloor \epsilon k \rfloor, \ldots, k \right\}$$

$$\Rightarrow \mathcal{M}_n^\epsilon := \left\{ \left( t^{-\xi} + \xi t^{-1} B(t), \frac{\xi}{1-\xi} t^{-\xi} + \xi t^{-1} \int_0^t y^{-(1+\xi)} B(y) dy \right) : \epsilon \leq t \leq 1 \right\} \text{ in } F, \quad (2.2)$$

where $B(t)$ is the standard Brownian bridge on $[0, 1]$. This is the case where $F$ has a finite second moment and hence the distributional limit has a Brownian component.
Case 2 \((1/2 < \xi < 1)\): For any \(0 < \varepsilon < 1\), as \(n, k, n/k \to \infty\),

\[
\mathcal{M}_n := \left\{ \left( \frac{X(i)}{k}, \frac{X(i)}{k} \right), \left( \frac{X(i)}{k}, \frac{X(i)}{k} \right) \right\} \Rightarrow \mathcal{M} := \left\{ \left( t^{-\xi} + \xi t^{-(1+\xi)} B(t), \frac{\xi}{1-\xi} t^{-\xi} + t^{-1} S_{1/\xi} \right), \varepsilon \leq t \leq 1 \right\} \text{ in } \mathcal{F},
\]

(2.3)

where \(S_{1/\xi}\) is a stable random variable with characteristic function

\[
E[e^{itS_{1/\xi}}] = \exp \left\{ -\frac{1}{1-\xi} \Gamma \left( 2 - \frac{1}{\xi} \right) \cos \frac{\pi}{2\xi} |t|^{1/\xi} \left[ 1 - i \text{sgn}(t) \tan \frac{\pi}{2\xi} \right] \right\},
\]

(2.4)

and is independent of the standard Brownian Bridge \(B(t)\) on \([0, 1]\). This is the case where \(F\) has a finite mean but does not have a finite second moment, hence we also observe a non-Gaussian stable weak limit. We need further conditions to find weak limits in the case where the tail parameter \(\xi = 1/2\), which we do not pursue here.

In the case where \(\xi \geq 1\), the ME function may not be well-defined. However, the ME plot can be defined for a dataset. Theorem 4.9 in Das and Ghosh (2013) describes the limit results of the ME plot when \(\xi \geq 1\).

2.2 ME plot in the Gumbel case

The behaviour (in probability) of ME plot when \(F\) is in the maximum domain of attraction of a Gumbel distribution has been discussed in Ghosh and Resnick (2010). We state the following result to recall notations to be used: this follows from Theorems 3.3.26 and 3.4.13(b) in Embrechts et al. (1997); see (Ghosh and Resnick 2010, Theorem 3.9) or (Resnick 2008, Proposition 1.4) or (de Haan and Ferreira 2006, Theorem 1.2.6) for further details.

**Proposition 2.1** The following are equivalent for a distribution function \(F\) with right end point \(x_F \leq \infty\):

1. \(F\) is in the maximum domain of attraction of the Gumbel distribution, i.e.,

\[
F^n \left( c(n)x + d(n) \right) \to \exp \left\{ -e^{-x} \right\} \text{ for all } x \in \mathbb{R},
\]

(2.5)

for some sequence \(c(n)\) and \(d(n)\).

2. There exists \(z < x_F\) such that \(F\) has a representation

\[
\tilde{F}(x) = \kappa(x) \exp \left\{ -\int_z^x \frac{1}{a(t)} dt \right\}, \text{ for all } z < x < x_F,
\]

(2.6)

where \(\kappa(x)\) is a measurable function satisfying \(\kappa(x) \to \kappa > 0, x \to x_F\), and \(a(x)\) is a positive, absolutely continuous function with density \(a'(x) \to 0\) as \(x \to x_F\).
We know from (Resnick 2008, Proposition 1.1) that a choice of the norming sequence \(c(n)\) and \(d(n)\) in (2.5) is
\[
d(n) = F^+ (1 - n^{-1}) \quad \text{and} \quad c(n) = a(d(n)).
\]
Theorem 3.3.26 in Embrechts et al. (1997) says that a choice of the auxiliary function \(a(x)\) in (2.6) is
\[
a(x) = \int_x^{x_F} \frac{F(t)}{F(x)} dt \quad \text{for all} \quad x < x_F,
\]
and for this choice, the auxiliary function is the ME function, i.e., \(a(x) = M(x)\). Furthermore, we also know that \(a'(x) \to 0\) as \(x \to x_F\) and this implies that \(M(u)/u \to 0\) as \(u \to x_F\). Define the ME plot in this case as
\[
\mathcal{M}_n := \left\{ \left( \frac{\ln i}{k} - X(i), \hat{M}(X(i)) \right) : i = 2, \ldots, k \right\}.
\] (2.7)
From a minor modification of (Ghosh and Resnick 2010, Theorem 3.10), we know that as \(n, k, n/k \to \infty\),
\[
\mathcal{M}_n \overset{P}{\to} \mathcal{M} := \{(t, 1) : t \geq 0\}.
\]
Now we will additionally put one more condition in order to get a weak limit for ME plots in the Gumbel case which is stated as follows.

**Assumption 2.2** The distribution function \(F\) satisfies the following:
\[
\sqrt{k} \left( \frac{\ln n}{k} F(c(n/k)y + d(n/k)) - e^{-y} \right) \to 0 \quad \text{(2.8)}
\]
point-wise and in \(L_1\)-norm in \([0, \infty)\) as \(n, k, n/k \to \infty\).

The assumption above is a second order condition to control the rate of convergence to the limit, and similar assumptions are quite usual in the extreme-value literature for proving weak convergence. Now we can state the distributional result for ME plots when \(F\) is in the maximum domain of attraction of the Gumbel distribution.

**Theorem 2.3** Suppose \(X_1, \ldots, X_n\) are i.i.d. observations from a distribution \(F\) which is in the maximum domain of attraction of the Gumbel distribution and satisfies Assumption 2.2. Then for any \(0 < \epsilon < 1\), as \(n, k, n/k \to \infty\),
\[
\mathcal{M}_n := \left\{ \left( -\ln \left( \frac{i}{k} \right), 1 \right) + \sqrt{k} \left( \frac{X(i) - X(k)}{X((i/k)\epsilon) - X(k)} + \ln \left( \frac{i}{k} \right), \frac{\hat{M}(X(i))}{X((i/k)\epsilon)} - 1 \right) : i = \lceil \epsilon k \rceil, \ldots, k \right\}
\]
\[
\Rightarrow \mathcal{M}_n := \left\{ \left( -\ln(t) + eB(e^{-1}) \ln(t) + \frac{B(t)}{t}, 1 + eB(e^{-1}) + \frac{1}{t} \int_0^t \frac{B(s)}{s} ds \right) , \epsilon \leq t \leq 1 \right\} \quad \text{in} \ F,
\]
where \(B(t)\) is a standard Brownian bridge on \([0, 1]\).

**Remark 2.4** As explained in Remark 2.4 in Das and Ghosh (2013), we look at the plot as the probability limit perturbed by the normalized deviation around it, i.e. we shift the normalized differences so that we can obtain the distribution of the deviation of the observed points of the ME plot from its mean position. Without this shift, we will not get the weak limit around the actual point in the plot.
Proof The proof is along the same lines of the proof of Theorem 4.3 in Das and Ghosh (2013). Hence we provide a sketch here, for a detailed argument along similar lines refer to the aforementioned theorem. Denote the tail empirical measures by

\[
\nu_n(\cdot) := \frac{1}{k} \sum_{i=1}^{n} \delta_{X_i - d(n/k)} \left( \frac{\cdot}{c(n/k)} \right) \quad \text{and,} \\
\hat{\nu}_n(\cdot) := \frac{1}{k} \sum_{i=1}^{n} \delta_{X_i - X(k)} \left( \frac{\cdot}{c(n/k)} \right)
\]

and define for \( k := k(n) < n \) and \( 0 < t \leq 1 \):

\[
W_n(t) := \sqrt{k} \left( \nu_n \left( -\ln t, \infty \right) - t, \nu_n \left( -\ln t, \infty \right) - t \right) = \sqrt{k} \left( \frac{1}{k} \sum_{i=1}^{n} \delta_{X_i - d(n/k)} \left( -\ln t, \infty \right) - t \right). \quad (2.11)
\]

We prove in Lemma 2.5 that \( W_n \Rightarrow W \) in \( D([0,1]) \), where \( W \) is the standard Brownian motion in \([0,1]\). Applying Vervaat’s lemma (Resnick 2007, Proposition 3.3, p.59) to (2.11) we get

\[
\sqrt{k} \left( \exp \left\{ \frac{-X(\lceil kt \rceil)}{c(n/k)} \right\} - t, \nu_n \left( -\ln t, \infty \right) - t \right) \Rightarrow (-W(t), W(t)) \quad \text{in} \ D^2([0,1]).
\]

Using the Functional Delta-method (van der Vaart and Wellner 1996, Theorem 3.9.4) we get

\[
\sqrt{k} \left( \frac{X(\lceil kt \rceil)}{c(n/k)} - \frac{d(n/k)}{c(n/k)} + \ln t, \nu_n \left( y, \infty \right) - e^{-y} \right) \Rightarrow \left( \frac{W(t)}{t}, W(e^{-y}) \right) \quad \text{in} \ D([0,1]) \times D([1,\infty]),
\]

and it is easy to check that

\[
\sqrt{k} \left( \frac{X(\lceil kt \rceil)}{c(n/k)} - \frac{d(n/k)}{c(n/k)} + \ln t, \hat{\nu}_n \left( y, \infty \right) - e^{-y} \right) \Rightarrow \left( \frac{B(t)}{t}, B(e^{-y}) \right) \quad \text{in} \ D([0,1]) \times D([1,\infty]),
\]

where \( B \) is a standard Brownian Bridge on \([0,1]\). Note aside that the above convergence also has the obvious implication, putting \( t = 1/e \) and restricting to the first component that

\[
\frac{X(\lceil k/e \rceil)}{c(n/k)} \Rightarrow -\ln 1/e = 1.
\]

Now following the same arguments used in the proof of Theorem 4.3 in Das and Ghosh (2013), where a continuous map is used judiciously we get

\[
\sqrt{k} \left( \frac{X(\lceil kt \rceil)}{c(n/k)} - \frac{X(k)}{c(n/k)} + \ln t, \hat{M}(X(\lceil kt \rceil)) \right) \quad \text{is} \quad \frac{k}{\lceil kt \rceil - 1} \int_{\nu_n(y, \infty) - 1}^{\infty} \hat{\nu}_n(y, \infty) dy - 1
\]

\[
\Rightarrow \left( \frac{B(t)}{t}, \frac{1}{t} \int_{-\ln t}^{\infty} B(e^{-y}) dy \right) = \left( \frac{B(t)}{t}, \frac{1}{t} \int_{0}^{t} \frac{B(s)}{s} ds \right) \quad \text{in} \ D^2([0,1]).
\]
We are able to replace \( c(n/k) \) above by \( X(\lceil k/e \rceil) - X(k) \) by using (2.14) and a Slutsky’s Theorem to obtain

\[
S_n(t) := \sqrt{k} \left( X(\lceil kt \rceil) - X(k) \right) + \ln t, \\
\hat{M}(X(\lceil kt \rceil)) - X(k) \]

\( \Rightarrow \) \( B(t) t \), \( \int_0^t B(s) ds \) = \( S(t) \) in \( D^2_l(0, 1) \).

The proof of the theorem is completed by using Lemma 2.4 in Das and Ghosh (2013) on \( S_n \) and \( S \).

Lemma 2.5 As \( n \to \infty, k \to \infty, n/k \to \infty, \)

\( W_n \Rightarrow W \) in \( D_l(0, 1) \) where \( W \) is a Brownian motion in \( D_l(0, 1) \).

Proof We check the conditions C1-C4 of (Rootzén 2009, Theorem 2.1). In this part of the proof whenever we write ‘\( \sim \)’ between two expressions, it means the asymptotics hold for \( n, k, n/k \to \infty \). Now following the notations used in the aforementioned paper, we set \( r_n = \min\{k^{1/4}, (n/k)^{1/2}\} \) and \( l_n = 1 \). For any \( u, v \in \mathbb{R} \) let

\[
N_n(u, v) := \sum_{i=1}^{r_n} \delta_{X_i - d(n/k)} (u, v). 
\]

Then for any \( \theta < x_T = \infty \) (since \( F \) is in a Gumbel domain of attraction, it has right end point \( x_T = \infty \)) with \( 0 \leq u, v < \theta \) we have,

\[
\mathbb{P}[N_n(u, v) \neq 0] \sim r_n \mathbb{P}[u c(n/k) + d(n/k) < X_1 \leq v c(n/k) + d(n/k)]
\]

and

\[
\mathbb{E}\left[N_n(u, v)^2 | N_n(u, v) \neq 0 \right] = \frac{\mathbb{E}\left[N_n(u, v)^2 \right]}{\mathbb{P}[N_n(u, v) \neq 0]} \sim 1 + r_n \mathbb{P}[u c(n/k) + d(n/k) < X_1 \leq v c(n/k) + d(n/k)] \\
\leq 1 + \text{const.} r_n k/n 
\]

which is bounded by the choice of \( r_n \). Hence condition C1 holds. Condition C2 holds as the random variables \( X_i \) are assumed to be independent. Next note that for any \( 0 \leq u, v < \infty, \)

\[
\frac{1}{r_n \hat{F}(d(n/k))} \text{Cov} \left( \sum_{i=1}^{r_n} \delta_{X_i - d(n/k)} (u, \infty), \sum_{i=1}^{r_n} \delta_{X_i - d(n/k)} (v, \infty) \right) = \frac{1}{\hat{F}(d(n/k))} \text{Cov} \left( \delta_{X_1 - d(n/k)} (u, \infty), \delta_{X_1 - d(n/k)} (v, \infty) \right) \sim \frac{\hat{F}((u \lor v)c(n/k) + d(n/k))}{\hat{F}(d(n/k))} \\
\to \exp(-u \lor v). 
\]
Hence C3 holds and obviously C4 holds because of the choice of \( r_n \). Hence, by (Rootzén 2009, Theorem 2.1)
\[
\sqrt{k} \left( v_n(u, \infty] - \mathbb{E}(v_n(u, \infty]) \right) \Rightarrow G \quad \text{in} \quad D[0, \infty),
\]
where \( G \) is a centered Gaussian process in \([0, \infty)\) with covariance function \( \exp(-u \vee v) \) and hence a time change \( u \mapsto -\ln u \) gives us that \( W_n \Rightarrow W \) in \( D_l(0, 1) \) where \( W \) is a standard Brownian Motion on \([0, 1]\).

### 2.3 ME plot in the Weibull case

If \( F \in D(G_{\xi, \beta}) \), then we have the following characterizations for the case \( \xi < 0 \) (Embrechts et al. 1997; Ghosh and Resnick 2010):

**Proposition 2.6** If \( \xi < 0 \) then the following are equivalent:

1. \( F \) has a finite right end point \( x_F \) and \( \bar{F}(x_F - x^{-1}) \in RV_{1/\xi} \).
2. \( F^n(x_F + c(n)x) \rightarrow \exp\{-(\gamma)^{-1/\xi}\} \) for all \( x \leq 0 \) where \( c(t) = x_F - F(1 - \frac{1}{t}) \), \( t \geq 1 \).
3. There exists a measurable function \( \beta(u) \) such that
   \[
   \lim_{u \to x_F} \sup_{u \leq x \leq x_F} |F_u(x) - G_{\xi, \beta(u)}(x)| = 0.
   \]

Recall from Ghosh and Resnick (2010), the following result on ME plots (there is a typographical error in the statement of the result there):

**Proposition 2.7** If \( X_1, \ldots, X_n \) are i.i.d. observations with distribution \( F \) which has a finite right end point \( x_F \) and satisfies \( 1 - F(x_F - x^{-1}) \in RV_{1/\xi} \) as \( x \to \infty \), then
\[
\nu_n := (\frac{1}{k_n X_{(1)} - X_{(k)}}) \left( (X_{(i)} - X_{(k)}, \tilde{M}(X_{(i)})) : i = 2, \ldots, k \right) \overset{p}{\to} \mathcal{M} := \left\{ (t, \frac{\xi}{1 - \xi} (1 - t)) : 0 \leq t \leq 1 \right\}.
\]
(2.15)

In this part of the paper we obtain the weak limit of the ME plot when the null hypothesis that \( \bar{F}(x_F - x^{-1}) \in RV_{1/\xi} \) for some \( \xi < 0 \) holds. In the same spirit as Das and Ghosh (2013) we deal with the tail empirical process. Denote by \( v_n \):
\[
v_n(\cdot) := \frac{1}{k} \sum_{i=1}^{n} \delta_{x_F - X_i}, \quad (2.16)
\]

Following Theorem 4.2 in Resnick (2007), we can show that
\[
v_n \Rightarrow v \quad \text{in} \quad M_+[0, \infty)
\]
where \( v[0, x] = x^{-1/\xi}, x \geq 0 \). Now define for \( k := k(n) < n \) and \( y \geq 0 \):
\[
W_n(y) := \sqrt{k} \left( \frac{1}{k} \sum_{i=1}^{n} \delta_{x_F - X_i}[0, y^{-\xi}] \right) - \frac{n}{k} \bar{F} \left( x_F - c(n/k)y^{-\xi} \right)
\]
\[
= \sqrt{k} \left( v_n[0, y^{-\xi}] - \mathbb{E}\left( v_n[0, y^{-\xi}] \right) \right).
\]
(2.17) (2.18)
The next result in the spirit of (Resnick 2007, Theorem 9.1) and also similar to Lemma 2.5.

**Lemma 2.8** As $n \to \infty$, $k \to \infty$, $n/k \to \infty$,

$$W_n \Rightarrow W$$

in $D[0, \infty)$ where $W$ is a Brownian motion in $D[0, \infty)$.

The proof follows by going through the steps of the proof of Lemma 2.5 or (Resnick 2007, Theorem 9.1). Let us also assume the following (again a second order type condition):

**Assumption 2.9** The distribution function $F$ satisfies the following

1. $\sqrt{k} \left( \frac{n}{k} \bar{F} (x_F - c(n/k)y) - y^{-1/\xi} \right) \to 0$ for all $y \geq 0$ locally uniformly,

2. $\sqrt{k} \int_0^1 \left| \frac{n}{k} \bar{F} (x_F - c(n/k)y) - y^{-1/\xi} \right| dy \to 0$ as $n, k, n/k \to \infty$.

**Theorem 2.10** Suppose $X_1, \ldots, X_n$ are i.i.d. observations from a distribution $F$ which has a finite right end point $x_F$ and satisfies $1 - F(x_F - x^{-1}) \in RV_{1/\xi}$, $\xi < 0$ as $x \to \infty$ and Assumption 2.9 holds. Then for any $0 < \epsilon < 1$, as $n, k, n/k \to \infty$,

$$\mathcal{M}_n := \left\{ \left( 1 - \left( \frac{i}{k} \right)^{-\xi} \cdot \frac{\xi}{\xi - 1} \left( \frac{i}{k} \right)^{-\xi} \right)^{i \in \epsilon k, \ldots, k} \right\} = \mathcal{M} := \left\{ \left( 1 - \epsilon^{-\xi} + \epsilon \xi t^{-1(1+\xi)} B(t) \right), \frac{\xi}{\xi - 1} \epsilon^{-\xi} + \epsilon \xi t^{-1} \int_0^t y^{-(1+\xi)} B(y) dy \right\}, \epsilon \leq t \leq 1 \right\}$$

in $F$, where $B(t)$ is the standard Brownian bridge on $[0, 1]$ restricted to $(0, 1]$.

**Remark 2.11** This result is similar to the one obtained for $\bar{F} \in RV_{-1/\xi}$ or $\xi > 0$ in Theorem 4.3 of Das and Ghosh (2013); the subtle difference appears in the fact that we no longer need to restrict the range of $\xi$ as is done there with $0 < \xi < 1/2$, since the integral

$$\int_0^t y^{-(1+\xi)} B(y) dy \overset{d}{=} \int_0^t y^{-(1+\xi)} W(y) dy - W(1) \int_0^t y^{-\xi} dy$$

exists if and only if $\int_0^t s^{-2\xi} ds < \infty$ which is always true for $\xi < 0$ and in turn implies that the limit $\mathcal{M}_n'$ exists. The truncation with $\epsilon$ with $\epsilon \leq t \leq 1$ is still necessary to guarantee that the limit set $\mathcal{M}_n'$ does not blow up for $t$ near $0$.

**Proof** The proof is omitted here as it follows using similar arguments as in the proof of (Das and Ghosh 2013, Theorem 4.3). The difference occurs in the fact that we use
the weak convergence result mentioned in Lemma 2.8 as our basis and apply a proper
version of Vervaat’s Lemma and ‘converging together’ arguments on this to obtain
the result.

3 Creating confidence bounds from the limit results

In Section 2 we obtain weak limits for ME plots for different values of \( \xi \in \mathbb{R} \) where
the underlying distribution \( F \in D(G_\xi) \). Now, depending on varying values of \( \xi \) we
construct the different confidence bounds following the results. We resort to Monte
Carlo simulation for actually computing the limits since most of them require calcu-
lating quantiles of suprema of functionals of Brownian bridges over a finite interval
or quantiles of stable distributions.

We need to truncate the ME plot near infinity in all the cases since the weak limits
we obtain blow up there (it relates to \( t \) near 0 in the limit of \( \mathcal{M}_n \)).

3.1 Fréchet case:

This case has already been discussed in Das and Ghosh (2013) and we recall it here
for the sake of completeness. Define the truncated versions of \( \mathcal{M}_n \) defined in (2.1)
and its limit \( \mathcal{M} \) respectively for \( 0 < \epsilon < 1 \) as:

\[
\mathcal{M}_n^\epsilon := \frac{1}{X(k)} \left\{ (X(i), \hat{M}(X(i))) : i = [k\epsilon], \ldots, k \right\}
\quad \text{and} \quad
\mathcal{M}^\epsilon := \left\{ \left( t^{-1}, \frac{\xi}{1 - \xi}, t^{-1} \right) : \epsilon \leq t \leq 1 \right\}. \tag{3.1}
\]

Then \( \mathcal{M}_n^\epsilon \xrightarrow{P} \mathcal{M}^\epsilon \).

Case 1 (0 < \( \xi < 1/2 \)): From (2.2), we have that the \((1 - \alpha)100 \%\) confidence band
for \( \mathcal{M}_n^\epsilon \) as

\[
\hat{C}\mathcal{M}_n^\epsilon := \mathcal{M}_n^\epsilon + \left\{ (x, y) : x \in \left( -\frac{c_{\alpha/2,\epsilon}}{\sqrt{k}}, \frac{c_{\alpha/2,\epsilon}}{\sqrt{k}} \right), y \in \left( -\frac{d_{\alpha/2,\epsilon}}{\sqrt{k}}, \frac{d_{\alpha/2,\epsilon}}{\sqrt{k}} \right) \right\}, \tag{3.2}
\]

where

\[
c_{\alpha,\epsilon} = (1 - \alpha)\text{-th quantile of } \sup_{\epsilon \leq t \leq 1} \xi t^{-(1+\xi)} B(t), \tag{3.3}
\]

\[
d_{\alpha,\epsilon} = (1 - \alpha)\text{-th quantile of } \sup_{\epsilon \leq t \leq 1} \xi t^{-1} \int_0^t y^{-(1+\xi)} B(y) dy. \tag{3.4}
\]

Since the weak limit of properly scaled and shifted \( \mathcal{M}_n^\epsilon \) consists of function-
als of the same Brownian Bridge in both components, (3.2) provides an asymptotic
confidence bound around \( \mathcal{M}^\epsilon \) with \( P(\mathcal{M}^\epsilon \subset \hat{C}\mathcal{M}_n^\epsilon) \geq (1 - \alpha) \) for large \( n \).
Case 2 \((1/2 < \xi < 1):\) From (2.3), we have the \((1 - \alpha)100\%\) confidence band for \(M^\epsilon\) as

\[
C M^\epsilon_n = \left\{ \left( \frac{X(\lceil k \epsilon \rceil)}{x(i)}, \frac{\hat{M}(X(\lceil k \epsilon \rceil))}{x(i)} \right) + \left(-\frac{c_{\alpha/2,\epsilon}}{\sqrt{k}}, \frac{d_{\alpha/2,\epsilon}}{\sqrt{k}}\right) \times \left( \frac{X(\lceil k \epsilon \rceil)}{\lceil k \epsilon \rceil}, \frac{X(\lceil k \epsilon \rceil)}{\lceil k \epsilon \rceil} \right) : \epsilon \leq t \leq 1 \right\},
\]

where

\[
d_{\alpha} = (1 - \alpha)-\text{th quantile of } S_{1/\xi} \text{ defined in (2.4)}.
\]

Here \(0 < \alpha_1, \alpha_2 < 1\) are chosen such that \((1 - \alpha) = (1 - \alpha_1)(1 - \alpha_2)\). Since the random components in the first and second components in the limit of (2.3) are independent this gives us the right confidence interval so that \(P(M^\epsilon C M^\epsilon_n) \geq 1 - \alpha\). The above quantiles are calculated using Monte Carlo simulation methods. In real data examples \(\xi\) is estimated using a Hill estimator, or any reasonable estimator for the tail index of a heavy-tailed distribution.

3.2 Gumbel case

This is the case where \(F \in D(G_0)\). Many well-known distribution functions such as exponential, normal, log-normal distributions fall into this class. First we define the truncated versions of \(M_n\) defined in (2.7) and its limit \(M\) respectively for \(0 < \epsilon < 1\) as:

\[
M^\epsilon_n := \frac{1}{X(\lceil k \epsilon \rceil) - X(k)} \left\{ X(i) - X(k), \hat{M}(X(i)) : i = [k \epsilon], \ldots, k \right\} \quad \text{and} \quad M^\epsilon := \{ (-\ln(t), 1) : \epsilon \leq t \leq 1 \}.
\]

Then \(M^\epsilon_n \overset{P}{\to} M^\epsilon\). Using Theorem 2.3, we have that the \((1 - \alpha)100\%\) confidence band for \(M^\epsilon\) as

\[
C M^\epsilon_n := M^\epsilon_n + \left\{ (x, y) : x \in \left( -\frac{c_{\alpha/2,\epsilon}}{\sqrt{k}}, \frac{d_{\alpha/2,\epsilon}}{\sqrt{k}} \right), y \in \left( -\frac{c_{\alpha/2,\epsilon}}{\sqrt{k}}, \frac{d_{\alpha/2,\epsilon}}{\sqrt{k}} \right) \right\}.
\]

where

\[
c_{\alpha,\epsilon} = (1 - \alpha)-\text{th quantile of } \sup_{\epsilon \leq t \leq 1} \left\{ eB(e^{-1}) \ln(t) + \frac{B(t)}{t} \right\}
\]

\[
d_{\alpha,\epsilon} = (1 - \alpha)-\text{th quantile of } \sup_{\epsilon \leq t \leq 1} \left\{ eB(e^{-1}) + \frac{1}{t} \int_0^t \frac{B(y)}{y} dy \right\}.
\]

By the same logic, as the earlier cases, (3.7) provides an asymptotic confidence bound around \(M^\epsilon\) with \(P(M^\epsilon \subset C M^\epsilon_n) \geq (1 - \alpha)\) for large \(n\). The quantiles are obtained using Monte Carlo simulation.

3.3 Weibull case

In this case \(F \in D(G_\xi)\) with \(\xi < 0\). Many distributions, especially with bounded right hand-tail falls into this category, for example Uniform, Beta, etc. Here we
define the truncated versions of $\mathcal{M}_n$ as defined in Proposition 2.7 and its limit $\mathcal{M}$ respectively for $0 < \epsilon < 1$ as:

$$\mathcal{M}_n^\epsilon := \frac{1}{X(1) - X(k)} \left\{ (X(i) - X(k), \hat{M}(X(i))) : i = [k\epsilon], \ldots, k \right\}$$

and $\mathcal{M}^\epsilon := \left\{ \left( t, \frac{\xi}{1 - \xi} (1 - t) \right) : \epsilon \leq t \leq 1 \right\}$.  

(3.10)

Then $\mathcal{M}_n^\epsilon \xrightarrow{P} \mathcal{M}^\epsilon$. Using Theorem 2.10, we have that the $(1 - \alpha)100\%$ confidence band for $\mathcal{M}^\epsilon$ as

$$\mathcal{C}\mathcal{M}_n^\epsilon := \mathcal{M}_n^\epsilon + \left\{ (x, y) : x \in \left( -\frac{c_{\alpha,\epsilon}}{\sqrt{k}}, \frac{c_{\alpha,\epsilon}}{\sqrt{k}} \right), y \in \left( -\frac{d_{\alpha,\epsilon}}{\sqrt{k}}, \frac{d_{\alpha,\epsilon}}{\sqrt{k}} \right) \right\},$$

(3.11)

where

$$c_{\alpha,\epsilon} = (1 - \alpha)\text{-th quantile of } \sup_{\epsilon \leq t \leq 1} \xi t^{-(1+\xi)} B(t),$$

(3.12)

$$d_{\alpha,\epsilon} = (1 - \alpha)\text{-th quantile of } \sup_{\epsilon \leq t \leq 1} \xi t^{-1} \int_0^t y^{-(1+\xi)} B(y) dy.$$

(3.13)

The bounds obtained here are very similar to the one in the Fréchet case. And using the same argument, (3.11) provides an asymptotic confidence bound around $\mathcal{M}^\epsilon$ with $P(\mathcal{M}^\epsilon \subset \mathcal{C}\mathcal{M}_n^\epsilon) \geq (1 - \alpha)$ for large $n$. Similar to the previous cases, the quantiles are obtained using Monte Carlo simulation.

**Remark 3.1** In statistical theory, while trying to obtain the confidence bound for a quantity, it is often imperative to obtain a prior point estimate of the quantity. That holds true in the case of obtaining confidence bounds for mean excess plots as well. The parametric form of the ME plot involves estimating $\xi$. Furthermore, the form of the confidence bound depends on the range of $\xi$, especially, whether $\xi > 0$, $\xi = 0$ or $\xi < 0$. We recommend choosing the null hypothesis regarding the maximal domain of attraction based on the estimate of $\xi$. If the estimate of $\xi$ is positive then we consider testing a null hypothesis of Gumbel domain of attraction with an alternate hypothesis of Fréchet domain of attraction. If the estimate of $\xi$ is negative then we consider testing a null hypothesis of Gumbel domain of attraction with an alternate hypothesis of Weibull domain of attraction.

**Remark 3.2** The choices of $k$ and $\epsilon$ remain with the user and is a hard problem without making assumptions of the rate of convergence to the max-domain of attraction. One should choose $k$ to be small enough to focus on the tail portion of the distribution, but big enough for the large sample behavior to take effect. The variance of the ME function explodes for the very top order statistics and one needs to remove that part from the calculation to obtain any meaningful bound for the ME plot. One should choose $\epsilon$ to not include the part where the data become sparse. In the simulation and real data studies described in Section 4 our choices of $k$ (5 % and 10 %) and
\(\epsilon (0.2 \text{ and } 0.3)\) were guided by these principles but is largely ad hoc. A data-driven approach to such choice is a subject of future research. A nice synopsis of methods to select \(k\) (and the threshold above which data can be considered exactly Generalized Pareto, which is a related problem) is available in Scarrott and MacDonald (2012); which we refer for further details on the topic.

4 Examples: simulated and real data

This section is devoted to application of the methodology developed for constructing confidence intervals around ME plots as derived in the Section 3.

Given an iid sample \(X_1, \ldots, X_n \sim F\), we are concerned with detecting if \(F \in D(G_\xi)\) and if so whether \(\xi\) is positive (the Fréchet case), zero (the Gumbel case) or negative (the Weibull case). The Fréchet case has been discussed in Das and Ghosh (2013) with examples. Hence we concern ourselves with the other two cases for the simulated examples. First we see how our confidence intervals work in simulated examples, and then use them on real data. In all the plots, the light blue shade creates a 99% confidence interval and the dark blue shade creates 90% confidence interval.

In the simulated examples the red dotted line corresponds to the ME plot for the known value of \(\xi\). In the real-data examples the red dotted line is obtained from the moment estimate of the tail parameter \(\xi\). In the Gumbel case the red-dashed line is always the line parallel to the horizontal axis at \(y = 1\).

4.1 Simulated examples: Weibull

In this case we have \(F \in D(G_\xi)\) with \(\xi < 0\). We consider two families of distributions here.

1. Consider \(F \sim \text{GPD}(\xi, \beta)\) with pdf given by

\[
F(x) = 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}, \quad 1 + \frac{\xi x}{\beta} > 0, \quad \beta > 0, \quad \xi < 0.
\]

Of course, \(F \in D(G_\xi)\) here and \(F\) is in the Weibull domain of attraction if \(\xi < 0\). In fact the Uniform \((0, 1)\) falls into this class with \(\xi = -1\) and \(\beta = 1\). For our simulation example we take \(\xi = -0.5, \beta = 1\) and generate 10000 iid samples from the distribution. The two plots in the first row of Fig. 1 are Pickands and Moment estimate of \(\xi\) for increasing values of top order statistics used. They seem reasonably close to \(-0.5\). For \(k = 800, 1000\) and \(\epsilon = 0.2, 0.3\) we create 90% (deep blue) and 99% (light blue) confidence bounds around the ME plot (in black) which clearly covers the dashed red line with slope \(-0.5\).

2. Next we consider \(F \sim \text{Beta}(a, b)\) with pdf given by

\[
f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1, \quad a, b > 0.
\]

In this case \(F \in D(G_{-1/b})\). We take the example \(a = 2, b = 2\) where \(F \in D(G_{-0.5})\). As also observed in the previous example, we see that in this example
too, the Pickands and Moment estimates approximate $-0.5$ well; see Fig. 2, first row. We again create confidence bounds with $k = 800, 1000$ and $\epsilon = 0.2, 0.3$ and observe that the bounds cover the dashed red line with slope $-0.5$ well.

Thus the detection in the Weibull family looks reasonable for these simulated data sets.

### 4.2 Simulated examples: Gumbel

Distributions in the Gumbel domain of attraction are harder to detect since the ME Plot from a data sample has to form a plot with slope zero in this case, which is statistically unlikely. Hence confidence bounds help to an extent, although as we will see through the three examples below that, in practice, a plotting technique is helpful to different degrees in different cases.

1. The first example is where $F$ follows $\text{Exp}(1)$. We generate 10000 iid samples from the distribution and create ME plots with parameters $k = 500, 1000$ and $\epsilon = 0.2, 0.3$; see Fig. 3. The Pickands and Moment estimates are close to zero and the confidence intervals around the ME plot in the four different cases all cover the line with slope 0 (and intercept 1) as expected. We also create Fréchet confidence bounds using the moment estimates of 0.03 (for $k = 1000$) and 0.07 (for $k = 500$), the bands are thin and doesn’t seem to cover the red-dashed line with slopes 0.03 and 0.07 in the respective cases. Hence we do not have evidence to claim that the true $\xi > 0$. In this example, both the Pickands and the Moment estimates are both close to 0 but the confidence bounds for the Gumbel case
definitely give us more confidence on the null hypothesis of $F \in D(G_0)$ which we cannot reject under the circumstances.

2. The next data sample is generated from $F$ which follows $N(0,1)$. We again generate 10000 iid samples from the distribution and make ME plots with parameters $k = 500, 1000$ and $\epsilon = 0.2, 0.3$; see Fig. 4. In this case both Pickands an Moment estimates are close to zero but they seem to be underestimates. The confidence intervals around the ME plot in the four different cases all cover the dashed red line with slope 0 (and intercept 1) up to some point. It appears to be covering the line better for $k = 500$. We cannot reject $F \in D(G_0)$ but the case is a little less convincing than the previous example. Using the moment estimates, the Weibull parameter is $\xi = -0.12$ (for $k = 1000$) and $\xi = -0.02$ (for $k = 500$). The confidence bounds are not that convincing especially for $k = 500$. For $k = 1000$, the confidence bounds are close to the dashed line, but not exactly covering it. Hence we have more confidence in the hypothesis that the underlying distribution is in the Gumbel maximum domain of attraction.

3. Finally we demonstrate simulation results for a standard lognormal distribution. The Lognormal distribution belongs $D(G_0)$ but is subexponential and has no finite moments (unlike the Normal or Exponential case). The Lognormal distribution is an ‘almost’ heavy-tailed distribution although it belongs to the Gumbel domain of attraction. We simulate 10000 iid samples from a standard Lognormal distribution and create ME plots as in the previous cases. The results are in Fig. 5. Both the Pickands estimate and the moment estimate of the extreme value parameter are much higher than the true value, that is zero. The ME plot with confidence intervals around it miss the target red dashed line of slope zero (and
Fig. 3 ME Plot for 10000 i.i.d. exponential random variables with $\lambda = 1$

intercept 1) clearly our technique doesn’t seem to perform so well here. In fact, a parameter estimate of $\xi = 0.3$ may be accepted as the true value of $\xi$ as indicated in the final figures of the plot. Since the Lognormal distribution has heavier tails than normal, exponential, or other distributions in the Gumbel domain we tend to have a positive slope of the ME plot as would happen in case when $F$ is in the Fréchet domain of attraction. Hence overall for detecting a Gumbel domain of attraction family one needs to be more careful with this technique.

4.3 Observed data: ozone concentration at Zurich urban area

It is of interest for environmental scientists to study ozone concentration near urban conglomerations, as its presence in the atmosphere implies health risks related to respiratory diseases. Directive 2008/50/EC of the European Parliament puts the target value of ozone for its member states to be within $120 \mu g/m^3$. The directive says that as of January 1, 2010 ozone concentrations should not exceed this limit for more than 25 days in a calendar year where the daily calculation is based on maximum of daily 8-hour averages.

We study a data set, freely available from www.eea.europa.eu. The data set contains daily maxima of ozone concentration (in $\mu g/m^3$) from one station in Zurich, Switzerland (station code CH 0010A, Zurich-Kaserne) located 410 mts above sea-level. Data is observed from January 1, 1992 to December 31, 2009. Measurements
were unavailable for 22 days, which we impute by the average value of ozone concentration on the same day for other available years.

As seen in the top left plot in Fig. 6 the data clearly admits periodicity. Moreover it is likely that the data is heteroscedastic. So we homoscedasticize the data by dividing the value on each date by the standard deviation of the values on the same day over all the 18 years of data available. Since our techniques work for stationary data sets, we fit an AR (38) process to the data set (AR(38) is chosen by an AIC criterion) and observe (from the ACF; see second plot from the second row in Fig. 6) that the residuals (first plot in the second row) look independent. Now we analyze the extremal behavior of the residuals of the model. The Pickands and Moments estimates provide a negative value but close to 0 and we can hypothesise that the sample is from a Weibull domain of attraction family. Since the value of the parameter is close to 0 we also check whether the data is possibly from a Gumbel domain of attraction family. The confidence bounds (90 % deep blue and 99 % light blue) are created assuming $F \in D(G_0)$ for $k = 327$ and $k = 654$; which are 5 % and 10 % of the data set and with $\epsilon = 0.2, 0.3$. From the plots in the third row we are unable to reject the null hypothesis that the maxima of the process belongs to a Gumbel domain of attraction with either 99 % or 90 % confidence bounds, although the estimate (black line) and the target (red-dashed line) do not touch at any point. This is most likely a result of the parameter being close to zero. The confidence bounds grow smaller as $\epsilon$ increases, so with a higher $\epsilon$, we would have perhaps rejected the null hypothesis in this case.
Fig. 5 ME Plot for 10000 i.i.d. standard log-normal random variables

On the other hand using the Moment estimator to estimate the tail index we get $\xi = -0.27$ for $k = 327$ and $\xi = 0.17$ for $k = 654$. The confidence bounds (again 90 % deep blue and 99 % light blue) for $k = 654$ in both cases $\epsilon = 0.2, 0.3$ covers the straight line with the slopes $\xi = -0.17$ quite well. On the other hand for $k = 327$, the coverage is not perfect but still close (see row 4, Fig. 6). Hence we expect that the underlying distribution is in fact in a Weibull domain of attraction with parameter close to $\xi = -0.2$.

4.4 Observed data: flow-rates at river Aare

The other data we analyze is maximum daily flow-rate at river Aare. River Aare flows through Switzerland and some manufacturing and power plants located near the river are often concerned about flooding on the river. The data we analyse has been collected from the Federal office of the Environment (FOEN), Bern and generously provided to us by Kernkraftwerk Gösgen-Däniken. It pertains to daily maximum flow-rates of Aare at the measurement station Aare-Murgenthal (2063) measured in $m^3/sec$ from 1st January 1974 to 20th October, 2010. See also www.hydrodaten.admin.ch/d/2063.htm.

Note that the data admits to possibilities of measurement error since automated measurement at the specific station started only in 1993. Moreover, the control authorities aim to maintain the flow-rate of Aare at the Aare-Murgenthal (2063) station below 850 $m^3/sec$ and would do so by using opening or closing log-gates. This
manually hinders the possibility of the data set being stationary. We were informed that such manual interventions have occurred a couple of times.

To analyse the data, we first note the seasonality pattern in the data set; see top left plot in Fig. 7. Hence as in the previous example we fit an AR process and work with the residuals obtained after the model-fitting (coincidentally an AR(38) is chosen by an AIC criterion here too). Observe that the Pickands and Moment estimates both indicate a small but positive value of the extreme value parameter; but does not completely reject the possibility of it being zero. We again create 90% (dark blue) and 99% (light blue) confidence bounds under the Gumbel assumption for $k = 671$ and $k = 1341$ (5% and 10%) of the sample size and $\epsilon = 0.2, 0.3$. The detection technique is unable to reject the null hypothesis that the underlying distribution $F \in D(G_0)$.

Now we allow a Moment estimator to chose the extreme value parameter which gives a value of around $\xi = 0.08$ (for $k = 670$) and $\xi = 0.07$ (for $k = 1340$). In both cases we create 90% and 99% confidence bounds with $\epsilon = 0.2, 0.3$. In all
the cases the confidence bounds are very narrow (barely visible in the plots) and do not cover the target red-dashed line at all. This seems to support that the data is from a distribution in the Gumbel domain of attraction, and not in a Fréchet domain of attraction.

5 Conclusion

A practitioner is often interested to understand the tail behavior of the underlying distribution of observations from a process in a data set. ME plots are extensively used as a visual diagnostic tool for this purpose since the slope of the ME function is positive, negative or zero depending on whether the underlying distribution can be justified to be a distribution in either a Fréchet (tail parameter $\xi > 0$), a Weibull ($\xi < 0$) or a Gumbel ($\xi = 0$) max-domain of attraction.
In practice we have little chance of observing a ME plot with zero slope. We frame this as the classic hypothesis testing problem where we are trying to test (i) $H_0 : \xi = 0$ vs. $H_1 : \xi \neq 0$ or (ii) $H_0 : \xi < 0$ vs. $H_1 : \xi \geq 0$, or, (iii) $H_0 : \xi > 0$ vs. $H_1 : \xi \leq 0$. In this paper we develop methods for obtaining confidence bounds for ME plots under the null hypothesis in consideration. The confidence bounds depend on $\xi$ and we use numerical estimates of $\xi$ from the data to compute the confidence intervals in practice.

In this paper we have concentrated on one plotting tool to detect tail behavior in data sets. There are other plotting tools in the heavy-tail and extreme value theory literature, like the Hill plot, probability plot, QQ plot, Gertensgarbe and Werner plot, Stărică plot, which are used as visual exploratory tools. The goal of this work is to bring about a framework to infer more about a data set in a rigorous manner using such plotting tools, and should provide groundwork for further work in this direction.

References

Beirlant, J., Goengebeur, Y., Teugels, J., Segers, J.: Statistics of extremes. Wiley series in probability and statistics. Wiley, Chichester (2004)

Billingsley, P.: Convergence of probability measures. Wiley, New York (1968)

Bingham, N.H., Goldie, C.M., Teugels, J.L.: Regular variation. Cambridge University Press (1987)

Das, B., Embrechts, P., Fasen, V.: Four theorems and a financial crisis. Int. J. Approx. Reason. 54, 701–716 (2013)

Das, B., Ghosh, S.: Weak limits of exploratory plots in the analysis of extremes. Bernoulli 19, 308–342 (2013)

Davison, A.C., Smith, R.L.: Models for exceedances over high thresholds (with discussion). J. R. Stat. Soc. Ser. B 52, 393–442 (1990)

de Haan, L.: On regular variation and its application to the weak convergence of sample extremes. Mathematisch Centrum, Amsterdam (1970)

de Haan, L., Ferreira, A.: Extreme value theory: an introduction. Springer, New York (2006)

Donnelly, C., Embrechts, P.: The devil is in the tails: actuarial mathematics and the subprime mortgage crisis. ASTIN Bulletin 40, 1–33 (2010)

Drees, H.: Extreme value analysis of actuarial risks: estimation and model validation. Advances in Statistical Analysis 96, 225–264 (2012)

Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling extreme events for insurance and finance. Springer, Berlin (1997)

Geluk, J.L., de Haan, L.: Regular variation, extensions and tauberian theorems. CWI Tract, vol. 40. Stichting Mathematisch Centrum, Centrum voor Wiskunde en Informatica, Amsterdam (1987)

Ghosh, S., Resnick, S.I.: A discussion on mean excess plots. Stoch. Process. Appl. 120, 1492–1517 (2010)

Hall, W.J., Wellner, J.: Mean residual life. In: Statistics and related topics (Ottawa, Ont., 1980, pp. 169–184. North-Holland, Amsterdam (1981)

Maulik, K., Resnick, S.I., Rootzén, H.: Asymptotic independence and a network traffic model. J. Appl. Probab. 39, 671–699 (2002)

McNeil, A.J., Frey, R., Embrechts, P.: Quantitative risk management: concepts, techniques and tools. Princeton University Press, Princeton (2005)

Neves, C., Fraga Alves, M.I.: Testing extreme value conditions—an overview and recent approaches. REVSTAT 6, 83–100 (2008)

Resnick, S.I.: Heavy tail phenomena: probabilistic and statistical modeling. Springer series in operations research and financial engineering. Springer, New York (2007)

Resnick, S.I.: Extreme values, regular variation and point processes. Springer series in operations research and financial engineering. Springer, New York (2008). Reprint of the 1987 original

Rootzén, H.: Weak convergence of the tail empirical process for dependent sequences. Stoch. Process. Appl. 119, 468–490 (2009)
Scarrott, C., MacDonald, A.: A review of extreme value threshold estimation and uncertainty quantification. REVSTAT 10, 33–60 (2012)

Seneta, E.: Regularly varying functions, p. 508. Springer, New York (1976). Lecture Notes in Mathematics

van der Vaart, A.W., Wellner, J.A.: Weak convergence and empirical processes with applications to statistics. Springer, New York (1996)

Yang, G.L.: Estimation of a biometric function. Ann. Stat. 6, 112–116 (1978)