Statistical Inference for inter-arrival times of extreme events in bursty time series

Abstract

In many complex systems studied in statistical physics, inter-arrival times between events such as solar flares, trades and neuron voltages follow a heavy-tailed distribution. The set of event times is fractal-like, being dense in some time windows and empty in others, a phenomenon which has been dubbed "bursty".

This article provides a new model for the threshold inter-exceedance *times*; the threshold exceedances are modeled via the standard Peaks Over Threshold (POT) method. For high thresholds and infinite-mean waiting times, we show that the times between threshold crossings are Mittag-Leffler distributed, and thus form a "fractional Poisson Process" which generalizes the standard Poisson Process of threshold exceedances. We provide graphical means of estimating model parameters and assessing model fit. Along the way, we apply our inference method to a real-world bursty time series, and show how the memory of the Mittag-Leffler distribution affects the predictive distribution for the time until the next extreme event.

Keywords: heavy tails renewal process extreme value theory peaks over threshold

1. Introduction

Time series displaying temporally inhomogeneous behaviour have received strong interest in the recent statistical physics literature (Barabási 2005; Oliveira and Barabási 2005; Vasquez et al. 2006; Vazquez et al. 2007; Omi and Shinomoto 2011; Min, Goh, and Vazquez 2011; Karsai et al. 2011; Bagrow and Brockmann 2013). They have been observed in the context of earthquakes, sunspots, neuronal activity and human communication (see Karsai et al. 2012; Vajna, Tóth, and Kertész 2013; Meerschaert and Stoev 2008 for a list of references). Such time series exhibit high activity in some 'bursty' intervals, which alternate with other, quiet intervals. Although several mechanisms are plausible explanations for bursty behaviour (most prominently self-exciting point process by Hawkes (1971)), there seems to be one salient feature which very typically indicates the departure from temporal homogeneity: a heavy-tailed distribution of waiting times (Vasquez et al. 2006; Karsai et al. 2012; Vajna, Tóth, and Kertész 2013). As we show below in simulations, a simple renewal process with heavy-tailed waiting times can capture this type of dynamics. For many systems, the renewal property is appropriate; a simple test of the absence of correlations in a succession of waiting times can be undertaken by randomly reshuffling the waiting times (Karsai et al. 2012).

Often a magnitude, or mark can be assigned to each event in the renewal process, such as for earthquakes, solar flares or neuron voltages. The Peaks-Over-Threshold model (POT, see e.g. Coles 2001) applies a threshold to the magnitudes, and fits a Generalized Pareto distribution to the threshold exceedances. A commonly made assumption in POT models is that times between events are either fixed or light-tailed, and this entails that the threshold crossing times form a Poisson process (Hsing, Hüsler, and Leadbetter 1988). Then as one increases the threshold and thus decreases the threshold crossing probability p, the Poisson process is rarefied, i.e. its intensity decreases *linearly* with p (see e.g. Beirlant et al. 2006).

As will be shown below, in the heavy-tailed waiting time scenario threshold crossing times form a *fractional Poisson process* (Laskin 2003; Meerschaert, Nane, and Vellaisamy 2011), which is a renewal process with Mittag-Leffler distributed waiting times. The family of Mittag-Leffler distributions nests the exponential distribution (Haubold, Mathai, and Saxena 2011), and hence the fractional Poisson process generalizes the standard Poisson process. Again as the threshold size increases and the threshold crossing probability p decreases, the fractional Poisson process is rarefied: The scale parameter of the Mittag-Leffler inter-arrival times of threshold crossing times increases, but *superlinearly*; see the Theorem below.

Maxima of events which occur according to a renewal process with heavy-tailed waiting times have been studied under the names "Continuous Time Random Maxima process" (CTRM) (Benson, Schumer, and Meerschaert 2007; Meerschaert and Stoev 2008; Hees and Scheffler 2016, 2017), "Max-Renewal process" (Silvestrov 2002; Silvestrov and Teugels 2004; Basrak and Špoljarić 2015), and "Shock process" (Esary and Marshall 1973; Shanthikumar and Sumita 1983, 1984, 1985; Anderson 1987; Gut and Hüsler 1999). The existing literature focuses on probabilistic results surrounding these models. In this work, however, we introduce a method of inference for this type of model, which is seemingly not available in the literature.

We review the marked renewal process in Section 2, and derive a scaling limit theorem for inter-exceedance times in Section 3. We give a statistical procedure to estimate model parameters via stability plots in Section 5, but to set the stage we first discuss inference for the Mittag-Leffler distribution in Section 4. In Section 6, we discuss the memory property of the Mittag-Leffler distribution, and how it affects the predictive distribution for the time until the next threshold crossing event. A simulation study of the effectiveness of our statistical procedure is given in Section 7. In Section 8 we apply our method to a real data set. Finally we close with a discussion and conclusion in Section 8. For all statistical computations we have used R (R Core Team 2018). All code and data used for the analysis in this article has been organized into an R package `CTRE` (https://github.com/UNSW-MATH/CTRE). The source code for the figures generated in this manuscript is available online at https://github.com/UNSW-MATH/bursty-POT.

2. Continuous Time Random Exceedances (CTRE)

In this section we will As a model for extreme observations, we use a Marked Renewal Process (MRP):

Definition (MRP): Let $(W, J), (W_1, J_1), (W_2, J_2), \ldots$ be i.i.d. pairs of random variables, where the $W_k > 0$ are interpreted as the *waiting times* and $J_k \in [x_L, x_R]$ as the *event magnitudes* ($x_L \in [-\infty, +\infty), x_R \in (-\infty, +\infty]$). If W and J are independent, the Marked Renewal Process is said to be *uncoupled*.

Note that the k-th magnitude J_k occurs at time $T_k = W_1 + \ldots + W_k$. Based on an MRP, we define the Continuous Time Random Exceedance model (CTRE) as follows:

Definition (CTRE): Given a threshold $\ell \in (x_L, x_R)$, consider the stopping time

$$\tau(\ell) := \min\{k : J_k > \ell\}, \quad \ell \in (x_L, x_R).$$

Define the pair of random variables $(X(\ell), T(\ell))$ via

$$X(\ell) = J_{\tau(\ell)} - \ell, \quad T(\ell) = \sum_{k=1}^{\tau(\ell)} W_k.$$

By restarting the MRP at $\tau(\ell)$, inductively define the two i.i.d. sequences $T(\ell, n)$ and $X(\ell, n)$, $n \in \mathbb{N}$, called the "interarrival times" and the "exceedances", respectively. The pair sequence $(T(\ell, n), W(\ell, n))_{n\in\mathbb{N}}$ is called a Continuous Time Random Exceedance model (CTRE). If the underlying MRP is uncoupled, then the CTRE is also called uncoupled.

In this article, we restrict ourselves to the uncoupled case, where W and J are independent. Then the two sequences $X(\ell, n)_{n\in\mathbb{N}}$ and $T(\ell, n)_{n\in\mathbb{N}}$ are independent as well. To see why, note that $X(\ell)$ is, in distribution, simply equal to $J|J > \ell$, independent of any waiting time W_k.
Figure 1 shows a simulated dataset in the top panel, where W has a stable distribution with tail parameter $\beta = 0.8$ (and skewness 1 and location 0), and where J is from a standard Gumbel distribution. In the bottom panel, we plot a time series of solar flare intensities derived from a NASA dataset (Dennis et al. 1991) which we will later examine more closely (see Section 8). Clearly, the simulated data exhibit long intervals *without any* events, whereas in the real-world dataset events appear continuously. The threshold exceedances, however, appear to have similar statistical behaviour in both models. Observations below a threshold are commonly discarded in Extreme Value Theory (POT approach); likewise, the CTRE model interprets these observations as noise and filters them out.

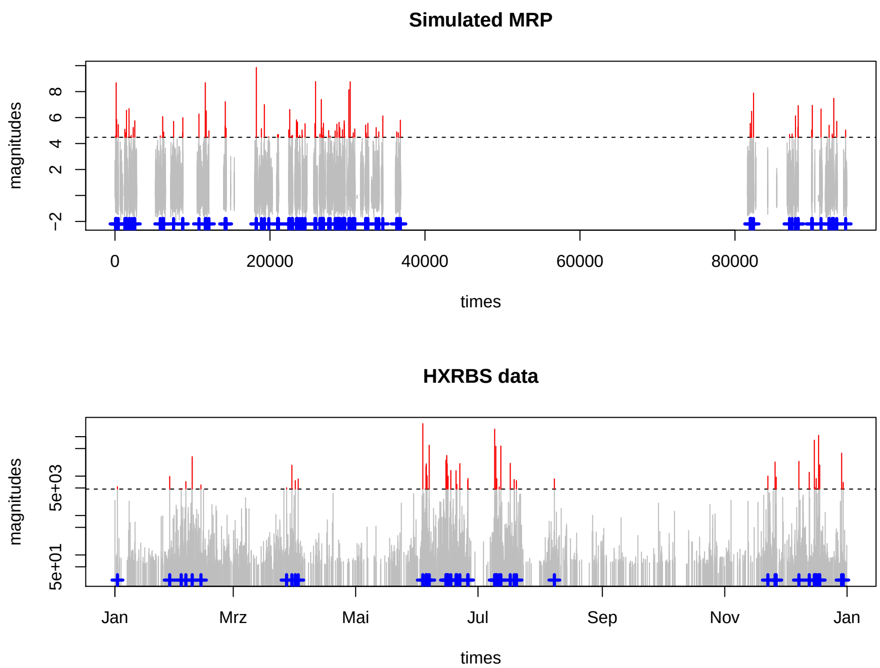

Figure 1: Exceedances (red) and Times until Exceedance (durations between blue crosses) for a given threshold ℓ (dashed line).

3. Scaling limit of Exceedance Times

In this section we state and prove the key theorem, see below. For an accessible introduction to regular variation and stable limit theorems, we recommend the book by Mark M Meerschaert and Sikorskii (2012).

Theorem: Let the waiting times J_k be in the domain of attraction of a positively skewed sum-stable law with stability parameter $0 < \beta < 1$; more precisely,

$$\frac{W_1 + \ldots + W_n}{b(n)} \xrightarrow{d} D, \quad n \to \infty \tag{1}$$

for a function $b(n)$ which is regularly varying at ∞ with parameter $1/\beta$, and where $\mathbf{E}[\exp(-sD)] = \exp(-s^\beta)$. Write $p := \mathbf{P}(J > \ell)$. Then the weak convergence

$$\frac{T(\ell)}{b(1/p)} \to W_\beta \quad \text{as} \quad \ell \uparrow x_R$$

holds, where the Mittag-Leffler random variable W_β is defined on the positive real numbers via

$$\mathbf{E}[\exp(-sW_\beta)] = \frac{1}{1 + s^\beta}.$$

Proof of Theorem: We interpret the threshold crossing time $T(\ell)$ as the hitting time of the underlying CTRM (Continuous Time Random Maxima) or "max-renewal process", and then utilize a result by Meerschaert and Stoev (2008). The running maximum process is defined as

$$M(c) := J_1 \vee \ldots \vee J_{\lfloor c \rfloor},$$

and since we assume that the J_k have a continuous distribution, there exist norming functions $a(c)$ and $d(c)$ such that

$$\mathbf{P}\left[\frac{M(c) - d(c)}{a(c)} \le \ell^*\right] \longrightarrow F(\ell^*), \quad t \to \infty$$

where F is a generalized extreme value distribution, and ℓ^* is any value from the support of F. The CTRM process is then defined via

$$V(t) = M(N(t)), \quad t \ge 0$$

where $N(t)$ is the renewal process associated with the waiting times W_k:

$$N(t) = \max\{n : W_1 + \ldots + W_n \le t\}.$$

Now a key observation is that

$$T(\ell) = \inf\{t : V(t) > \ell\},$$

and that

$$T(\ell) > t \quad \text{if and only if} \quad V(t) \leq \ell.$$

By (Theorem 3.1, Meerschaert and Stoev 2008), we have the stochastic process convergence

$$\frac{V(ct) - d(\tilde{b}(c))}{a(\tilde{b}(c))} \xrightarrow{d} Y(t), \quad t > 0.$$

where $Y(t)$ is a time-changed ("subordinated") extremal process, and where $\tilde{b}(c)$ is a regularly varying norming function which is *inverse* to $b(c)$, in the sense that $b(\tilde{b}(c)) \sim c \sim \tilde{b}(b(c))$.

Without loss of generality, we choose ℓ^* such that $F(\ell^*) = 1/e$, and let $\ell = a(\tilde{b}(c))\ell^* + d(\tilde{b}(c))$. We may then calculate

$$\mathbf{P}\left[\frac{T(\ell)}{b(1/p)} > t\right] = \mathbf{P}[T(\ell) > b(1/p)t] = \mathbf{P}[V(ct) \leq \ell]$$

where we have substituted $c = b(1/p)$. Moreover

$$\mathbf{P}[V(ct) \leq \ell] = \mathbf{P}\left[\frac{V(ct) - d(\tilde{b}(c))}{a(\tilde{b}(c))} \leq \frac{\ell - d(\tilde{b}(c))}{a(\tilde{b}(c))}\right] \longrightarrow \mathbf{P}\left[Y(t) \leq \ell^*\right]$$

Defining the hitting time of level ℓ^* by $Y(t)$ as $\xi_{\ell^*} := \inf\{t : Y(t) > \ell^*\}$, we then have

$$P\left[Y(t) \leq \ell^*\right] = \mathbf{P}[\xi_{\ell^*} > t] = \mathbf{P}[(-\log F(\ell^*))^{-1/\beta} X^{1/\beta} D > t]$$

by (Proposition 4.2, Meerschaert and Stoev 2008), where X is an exponential random variable with mean 1. Using (Theorem 19.1, Haubold, Mathai, and Saxena 2011), we see that $X^{1/\beta} D \sim \mathrm{ML}(\beta, 1)$, concluding the proof. □

For a scale parameter $\sigma > 0$, we write $\mathrm{ML}(\beta, \sigma)$ for the distribution of σW_β. The Mittag-Leffler distribution with parameter $\beta \in (0, 1]$ is a heavy-tailed positive distribution for $\beta < 1$, with infinite mean. However, as $\beta \uparrow 1$, $\mathrm{ML}(\beta, \sigma)$ converges weakly to the exponential distribution $\mathrm{Exp}(\sigma)$. This means that although its moments are all infinite, the Mittag-Leffler distribution may (if β is close to 1) be indistinguishable from the exponential distribution, for the purposes of applied statistics. For a detailed reference on the Mittag-Leffler distribution, see e.g. Haubold, Mathai, and Saxena (2011), and for algorithms, see e.g. the R package `MittagLeffleR` (Gill and Straka 2017).

Remark: If $\beta = 1$, the result of the Theorem above is standard, see e.g. Equation (2.2) in Gut and Hüsler (1999). In Anderson (1987) a similar result is shown with a different choice of scaling constant.

4. Model Choice and inference for the Mittag-Leffler distribution

The classical peaks over threshold approach is based on the fact, that exceedances above a high threshhold are assymptotically GPD distributed. Hence, a GPD distribution is fitted to the exceedances above a high enough threshhold.

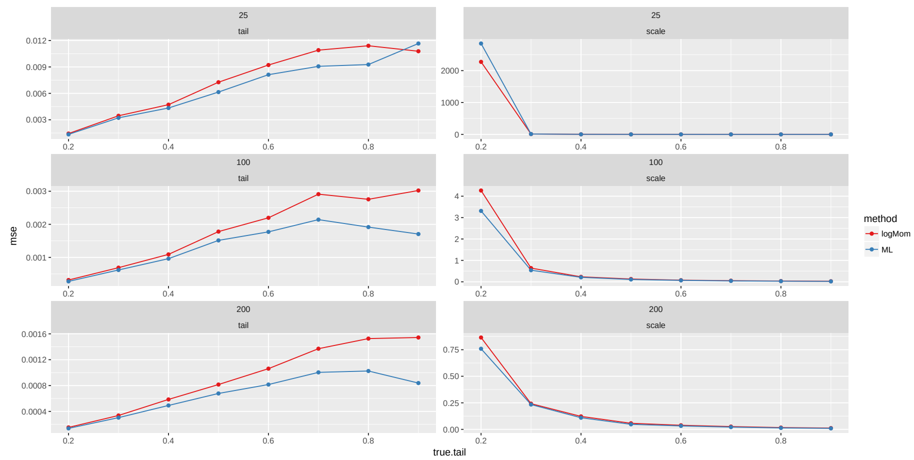

Figure 2: MSE for sample sizes n=30, 100, 200 for the tail and scale estimations via log-moments and maximum likelihood estimation

Due to the Theorem in the last Section, we know that for high threshholds the times between the exceedances are in case of heavy tailed inter-arrival times assymptotically Mittag-Leffler distributed. Therefore, before talking about inference for the exceedance times in the next section, we want to have a look at inference for Mittag-Leffler distributions in general.

The first method proposed in the literature to estimate the parameters of Mittag-Leffler distribution was the fractional moment estimator by (Kozubowski 2001). Unlike the first moments, the fractional moments of order p for $p < \beta$ exist and can get calculated by explicit formulas. One drawback of this method is that constant priors for the tail parameter are needed for the calculation of the estimates. Therefore, (Cahoy, Uchaikin, and Woyczynski Wojbor 2010) proposed an estimator based on the log-Moments which doesn't need prior knowledge. Furthermore, they showed in a simulation study that the log-Moment outclasses the fractional moment estimator regarding bias and root mean squared error (RMSE).
Due to the form of the density function of a Mittag-Leffler distribution there exists no closed form for the Maximum Likelihood estimator (MLE). In the RPackage `MittagLeffleR` we implemented Maximum Likelihood estimation via an numerical optimization algorithm. The MLE outperforms the log-moment estimator regarding bias and RMSE for big enough sample sizes, but is extremely computational intensive and the differences in bias and RMSE between the two methods are very small. In Figure 2, one can see that both methods work quite well even for very low sample sizes.

Since the Mittag-Leffler distribution is heavy-tailed, many researchers would intuitively give the highest importance to the tail behaviour of the distribution. Of course, one can also use established tail exponent estimators for the estimation

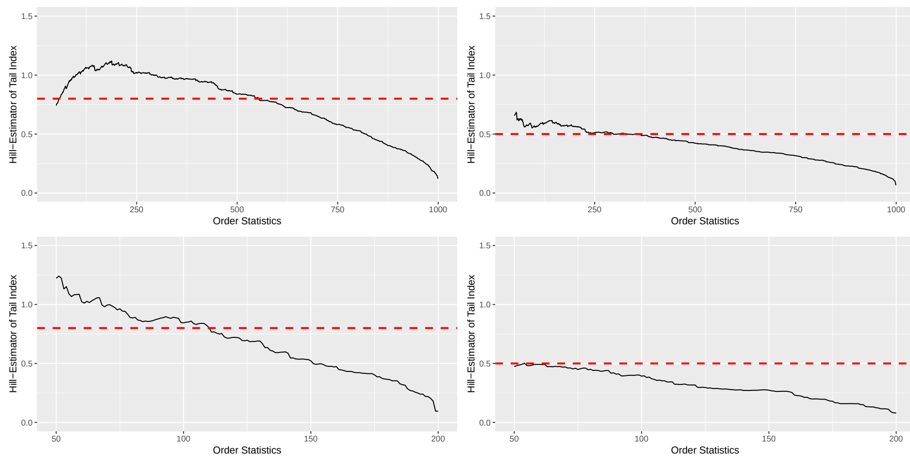

Figure 3: Hillplots for simulated Mittag-Leffler distributed rv's with tail parameter 0.5 (two plots in the right column) and 0.8 (two plots in the left column) and sample sizes 200 (two plots in the lower row) and 1000 (two plots in the upper row).

of the parameter β such as the Hill estimator, which is defined by

$$H_{k,n} = \frac{1}{k} \sum_{i=0}^{k-1} \log \frac{X_{n-i,n}}{X_{n-k,n}}.$$

However, these methods are less statistically efficient since they only use a portion of the information contained in the data, as was already mentioned by (Kozubowski 2001). Additionally, they need a specification of a tuning parameter and the hill estimator only performs well for distributions close to Pareto (see Resnick (1997)). In Figure 3 one can see different Hill plots for Mittag-Leffler simulated data with different sample sizes and tails. It would be really hard to deduce the correct tail parameter estimates from these plots.

Since the exponential distribution is nested in the Mittag-Leffler family of distributions, an appropriate way to choose between a model with exponential and Mittag-Leffler inter-exceedance times seems to be a Likelihood ratio test. Although the two models are nested, the assymptotic distribution is not χ^2_1-distributed, the classical Theorem of Wilk doesn't apply since under H_0 the parameter β of the Mittag-Leffler distribution is equal to one and hence lies under H_0 on the boundary of the parameter space $(0, 1]$. Nonetheless, one can perform a bootstrapped log-likelihood ratio test. In Figure 4 one can see the power for the bootstrapped LRT for Mittag-Leffler distributions with different tail parameters. As expected the power gets smaller for tail parameters close to one, since the Mittag-Leffler distribution converges for $\beta \to 1$ to an exponential distribution. Consequently, for tail parameters close to one it is hard to differentiate a Mittag-Lefller distribution from an exponential.

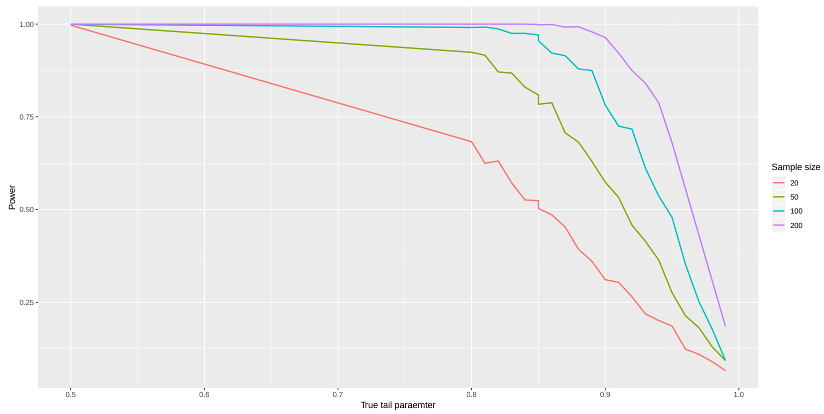

Figure 4: Power for bootstrapped LRT for different sample sizes

5. Inference on Exceedance times

The Theorem in Section 3 implies that for a high threshold ℓ we may approximate the distribution of $T(\ell)$ with an $\mathrm{ML}(\beta, b(1/p))$ distribution, where the function $b(c)$ varies regularly at ∞ with parameter $1/\beta$. Building on the POT (Peaks Over Threshold) method, we propose the following estimation procedure for the distribution of inter-exceedance time $T(\ell)$:

1. For a range of thresholds ℓ near the largest order statistics, extract datasets of exceedance times $\{T(\ell, i)\}_i$.
2. For each choice of threshold ℓ, fit a Mittag-Leffler distribution to the resulting dataset $\{T(\ell, i)\}_i$. This results in the estimates $\{\hat\beta(\ell)\}_\ell$ and $\{\hat\sigma(\ell)\}_\ell$.
3. Plot ℓ vs. $\hat\beta(\ell)$. As ℓ increases towards x_R, $\hat\beta(\ell)$ *stabilizes* around a constant $\hat\beta$. Use $\hat\beta$ as an estimate for the tail parameter β of the Mittag-Leffler distribution of exceedance times.
4. Approximate $p \approx |\{k : J_k > \ell\}|/n$. Recall that $b(c)$ is regularly varying with parameter $1/\beta$, and hence has the representation $b(c) = L(c)c^{1/\beta}$ for some slowly varying function $L(c)$. Assuming that the variation of $L(c)$ is negligible, we hence plot ℓ vs. $p^{1/\hat\beta}\hat\sigma(\ell)$. Again as ℓ increases towards x_R, $p^{1/\hat\beta}\hat\sigma(\ell)$ is expected to stabilize around a constant $\hat\sigma_0$. We then use $p^{-1/\hat\beta}\hat\sigma_0$ as an estimate of the scale parameter of the Mittag-Leffler distribution of exceedance times for the level ℓ.

The above approach, though theoretically sound, benefits from the following practical adjustments (compare with Figure 11):

- We choose ℓ from the order statistics, i.e. ℓ is the k-th largest of the observations X_j, where k runs from $k_{\min}, k_{\min}+1, \ldots, k_{\max}$. The datasets are then of length $k-1$.

- We use k rather than ℓ for the horizontal axis of our plots.
- In Step 4, rather than plotting $p^{1/\hat{\beta}}\hat{\sigma}(\ell)$ we plot $k^{1/\hat{\beta}}\hat{\sigma}(\ell)$. This changes $\hat{\sigma}_0$ by the multiplicative constant $n^{1/\hat{\beta}}$, but has the advantage that $\hat{\sigma}_0$ does not change if one pre-processes the data by removing all observations below a certain threshold.

The estimates $\hat{\beta}$ and $\hat{\sigma}_0$ give an estimate of the distribution of exceedance times, dependent on the threshold ℓ:

$$T(\ell) \sim \mathrm{ML}(\hat{\beta}, k^{-1/\hat{\beta}}\hat{\sigma}_0).$$

For quick estimates of the Mittag-Leffler parameters we have used the method of log-transformed moments by Cahoy (2013). We have verified the validity of our estimation algorithm via simulations, see the appendix.

6. Predicting the time of the next threshold crossing

According to Figure 11, for a threshold ℓ at the k-th order statistic, the fitted threshold exceedance time distribution is

$$T(\ell) \sim \mathrm{ML}(\beta, k^{-1/\beta}\sigma_0),$$

where $\beta = 0.85$ and $\sigma_0 = 3.0 \times 10^7$sec. Unlike the exponential distribution, the Mittag-Leffler distribution is not memoryless, and the probability density of the time t until the next threshold crossing will depend on the time t_0 elapsed since the last threshold crossing. This density equals

$$p(t|\beta, \sigma_0, \ell, t_0) = \frac{f(t + t_0|\beta, k^{-1/\beta}\sigma_0)}{\mathbf{P}[T_\ell > t_0]}$$

where $f(\,\cdot\,|\beta, k^{-1/\beta}\sigma_0)$ is the probability density of $\mathrm{ML}(\beta, -k^{1/\beta}\sigma_0)$. The more time has passed without a threshold crossing, the more the probability distribution shifts towards larger values for the next crossing (see Figure 5, left panel). The hazard rate

$$h(t) = \frac{f(t|\beta, k^{-1/\beta}\sigma_0))}{\int_t^\infty f(\tau|\beta, k^{-1/\beta}\sigma_0))\, d\tau}$$

represents the risk of a threshold crossing per unit time, and is a decreasing function for the Mittag-Leffler distribution.

The closer β is to 1, the more the hazard rate mimics that of an exponential distribution (a constant function, see Figure 5, right panel).

7. Simulation study

To test our inference method, we performed a simulation study where we have simulated independent waiting time and magnitude pairs (W_k, J_k). Since

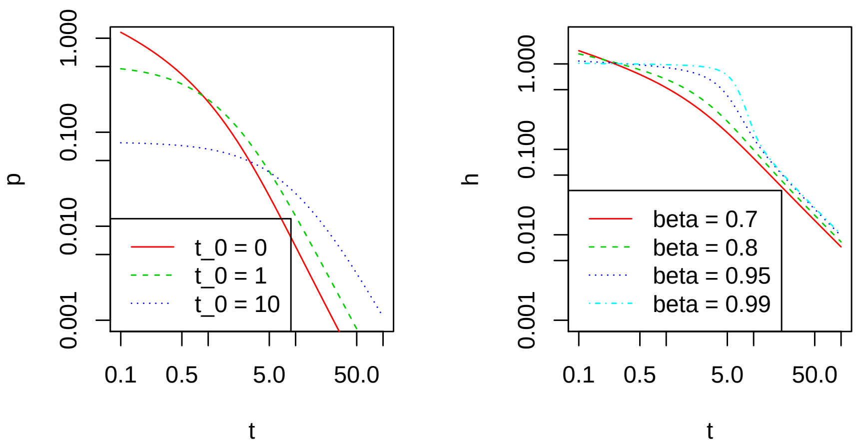

Figure 5: Left: Conditional distribution of time until next threshold crossing, depending on elapsed time t_0 since last crossing ($\beta = 0.8$, $\sigma_0 = 1$). Right: Hazard rate depending on tail parameter β.

the estimation of the parameters of the inter-exceedance times is based on stability plots and is hence rather a graphical estimation procedure, it is difficult to conduct a simulation study. To get an impression of how our estimation procedure performs, we plotted $m = 100$ stability plots for each example and calculated for each k the mean value of all estimates. We have chosen (i) a Stable, (ii) a Pareto and (iii) an Exponential distribution for waiting times. The distribution of J_k is irrelevant for the inference on β and σ_0 (it just plays a role for the rescaling of $\hat{\sigma}_0$, which gets done to receive an estimate for the scale parameter of the exceedances). Hence we have chosen exponential distributed magnitudes in all considered examples.
In order to have exact analytical values available for β and σ_0, a distribution for W_k needs to be chosen for which $b(n)$ from (1) is known. In our first example, we have chosen $W_k \stackrel{d}{=} D$, where D is as in (1). Then due to the stability property we have the *equality* of distribution $W_1 + \ldots + W_n \stackrel{d}{=} b(n)D$, for $b(n) = n^{1/\beta}$. Using the parametrisation of Samorodnitsky and Taqqu (1994), a few lines of calculation (see e.g. the vignette on parametrisation in Gill and Straka (2017)) show that D must have the stable distribution $S_\beta(\cos(\pi\beta/2)^{1/\beta}, +1, 0)$, which is implemented in the R package `stabledist` by Wuertz, Maechler, and members. (2016). In the second examaple we have chosen W_k to be Pareto distributed or more precisecly

$$P(W_k > t) = Ct^{-\beta}$$

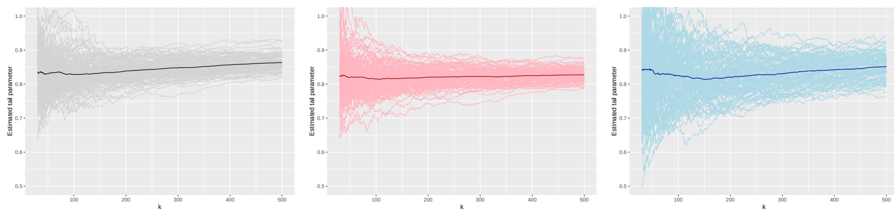

Figure 6: Stability Plots for m=100 simulation runs for Example 1 (stable distributed waiting times, true tail parameter =0.8). The grey lines, the red line and the blue lines are the log Moment, the MLE and the hill estimator, repectively, for different k's on the x-axis. The dark lines in each picture are the means of the concerning estimators for different k's.

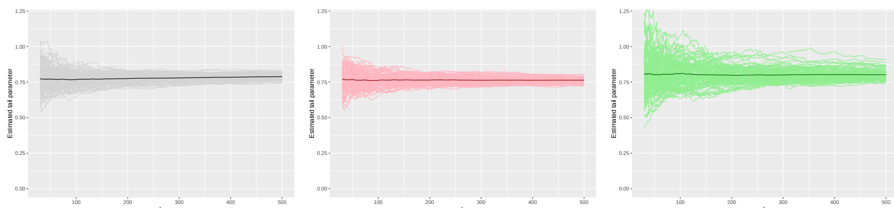

Figure 7: Stability Plots for m=100 simulation runs for Example 2 (Pareto distributed waiting times, true tail parameter =0.8). The grey lines, the red line and the blue lines are the log Moment, the MLE and the hill estimator, repectively, for different k's on the x-axis. The dark lines in each picture are the means of the concerning estimators for different k's.

with $C := (1/\Gamma(1-\beta))^{1/\beta}$ and $b(n) = n^{1/\beta}$. We then receive

$$\frac{W_1 + \ldots + W_n}{b(n)} \overset{d}{\Rightarrow} D$$

with $E(\exp(-sD)) = \exp(-s^{\beta})$.

By the Theorem, the distribution of $T(\ell)$ is approximately

$$\mathrm{ML}(\beta, p^{-1/\beta}) = \mathrm{ML}(\beta, k^{-1/\beta} n^{1/\beta}),$$

which means $\sigma_0 = n^{1/\beta}$. we have chosen $\beta = 1$ in the stable as well in the Pareto case. Figure **??** displays plots of $\hat{\beta}(k)$ vs. k for the first example with stable distributed waiting times. In the left picture we used the log moment estimator, in the middle picture maximum likelihood estimates (both applied to the inter-exceedance times) and for the right picture hill estimators applied to the inter-arrival times. Recall that k is the index of the order statistics of J_k at which the threshold ℓ is placed.

Figure **??** shows the estimates fordifferent k's and for the different estimators for the tail parameter β. Again one can see, that there is much more variance between the stability plots for the different simulation runs in case of the hill estimator compared to the other two estimators.

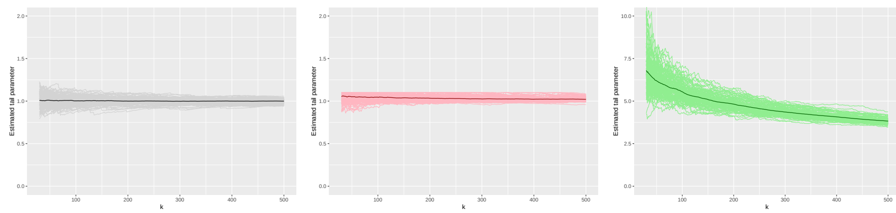

Figure 8: Stability Plots for m=100 simulation runs for Example 3 (Exponentially distributed waiting times, true tail parameter =0.8). The grey lines, the red line and the blue lines are the log Moment, the MLE and the hill estimator, repectively, for different k's on the x-axis. The dark lines in each picture are the means of the concerning estimators for different k's.

To show how the different estimates behave in the case that $\beta = 1$, hence, the inter-exceedance times are exponetial distributed, we simulated in a last example exponetial distributed waiting times.

8. Data example

We now want to apply the proposed method to a real data example, the solar flare data which was already mentioned in Section 1 and can be seen in Figure 1. The data were extracted from the "complete Hard X Ray Burst Spectrometer event list", a comprehensive reference for all measurements of the Hard X Ray Burst Spectrometer on NASA's Solar Maximum Mission from the time of launch on Feb 14, 1980 to the end of the mission in Dec 1989. 12,776 events were detected, with the "vast majority being solar flares". The list includes the start time, peak time, duration, and peak rate of each event. We have used "start time" as the variable for event times, and "peak rate" as the variable for event magnitudes.

Before we apply the POT approach described in Section 5 to the solar flare data, we first have to check if all model assumptions are full-filled. The CTRE model is based on three main assumptions, which are repeated below. For each assumption, we suggest one means of checking if it holds:

i.i.d.: After removing the "noise observations" below the smallest threshold ℓ_0, the pair sequence $(T(\ell_0, i), X(\ell_0, i))$ is i.i.d. An indication if this is true is given by an auto-correlation plot for the logarithms (to ensure finite moments) of the two time series.

Uncoupled: Each $T(\ell, i)$ is independent of each $X(\ell, i)$. We propose an empirical copula plot to check for any dependence.

ML(β, σ) **distribution of** $T(\ell, i)$**:** Apply a cutoff at the lowest threshold ℓ_0, extract the threshold crossing times, and create a QQ Plot for the Mittag-Leffler distribution. Use a log-Moment estimate of the tail parameter for the theoretical / population quantiles of the plot.

Figures 9 and 10 show the diagnostic plots for a minimum threshold chosen at the 200th order statistic. There is some residual autocorrelation for the sequence of threshold exceedance times that is not accounted for by the CTRE model.

Figure 11 shows the stability plots for the solar flare data, on the left for the tail parameter and on the right for the scale parameter. Dotted lines show 95% confidence intervals, which are derived from the asymptotic normality of the log-moments estimators (Cahoy 2013) and the δ-method [Gill and Straka (2017), dashed lines show the actual values of β resp. σ_0. The stability plot for the tail stabilizes nicely round about 0.85 (dotted line), while it is a little bit harder to deduce an estimate for the scale parameter. Howerever, the stability plot for the scale parameter seems to stabilize round about 3×10^7.

The fit with a Mittag-Leffler distribution ($\beta = 0.85$) is good, though there are signs that the power-law tail tapers off for very large inter-threshold crossing times. There is no apparent dependence between threshold exceedance times and event magnitudes seen in the copula plot. We also counduct a bootstrapped LRT for the null hypothesis of exponential distributed inter-arrival times and received a p-value of $p < 0.01$.

9. Discussion & Conclusion

We have extended the POT (Peaks over Threshold) model, a mainstay of extreme value theory, to "bursty" time series, which have been studied intensively in statistical physics. Burstiness is characterized by power-law waiting times between events, and we have shown that the Mittag-Leffler distribution arises naturally as a scaling limit for the inter-exceedance times of high thresholds. Moreover, we have derived the following non-linear scaling behaviour: $\sigma \sim p^{-1/\beta}$, where σ is the scale parameter of the distribution of threshold exceedance times, p is the fraction of magnitudes above the threshold, and β the exponent of the power law. This "anomalous" scaling behaviour in the bursty setting entails two phenomena:

i) a heavy tail of the interarrival time distribution of threshold crossings (long rests), and
ii) a high propensity for more threshold crossing events immediately after each threshold crossing event (bursts).

The Mittag-Leffler distribution captures both phenomena, due to its heavy tail as well as its stretched exponential (peaked) asymptotics for small times. It generalizes the exponential distribution, and in the solar flare data example, this generalization is warranted, because the likelihood-ratio test is strongly significant.

When we introduced the CTRE model, we assumed that all events are i.i.d. This assumption is likely sufficient but not necessary for our limit theorem to hold. Moreover, any data below a (minimum) threshold ℓ_0 is discarded for CTREs, and hence need not satisfy the i.i.d. assumption.For the purposes of statistical inference, we merely require that the inter-threshold-crossing times are i.i.d.

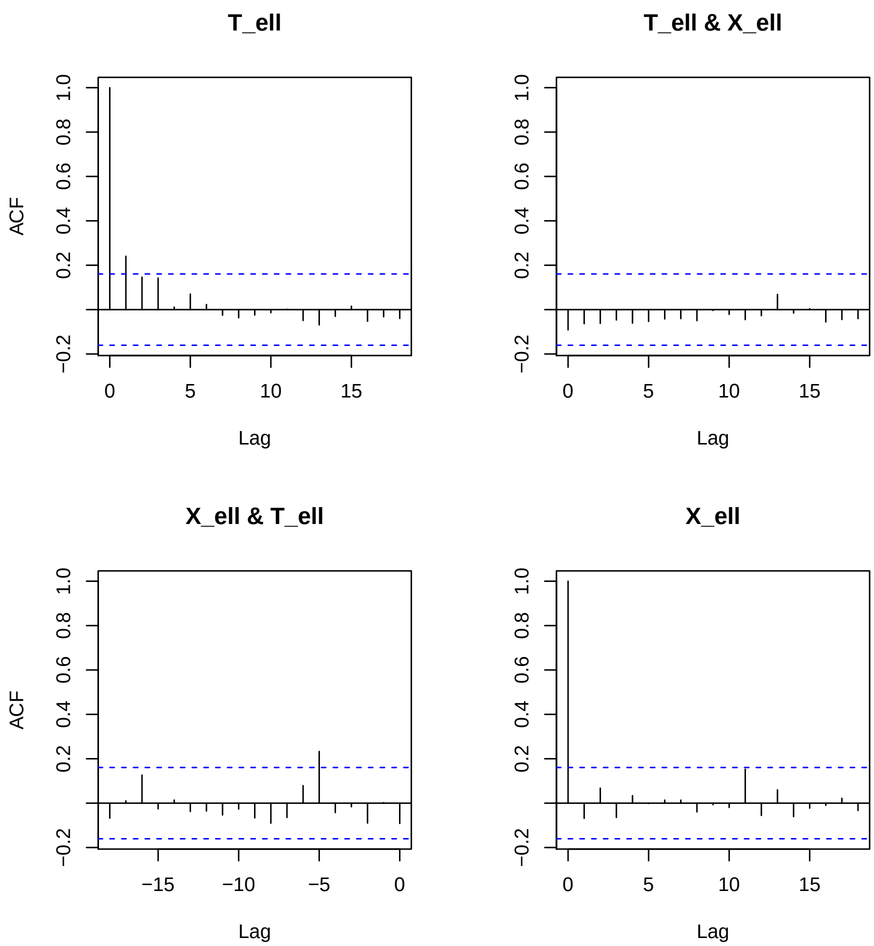

Figure 9: Diagnostic plots for the solar flare data: auto-correlation function.

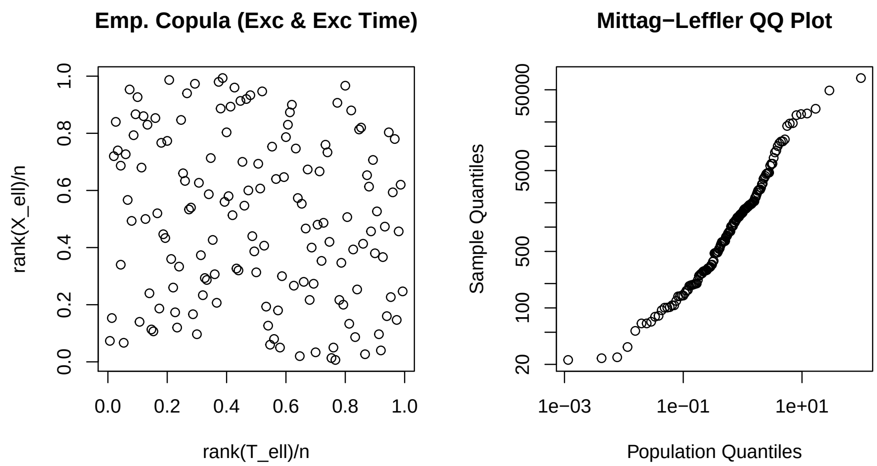

Figure 10: Diagnostic plots for the solar flare data: empirical copula and QQ Plot.

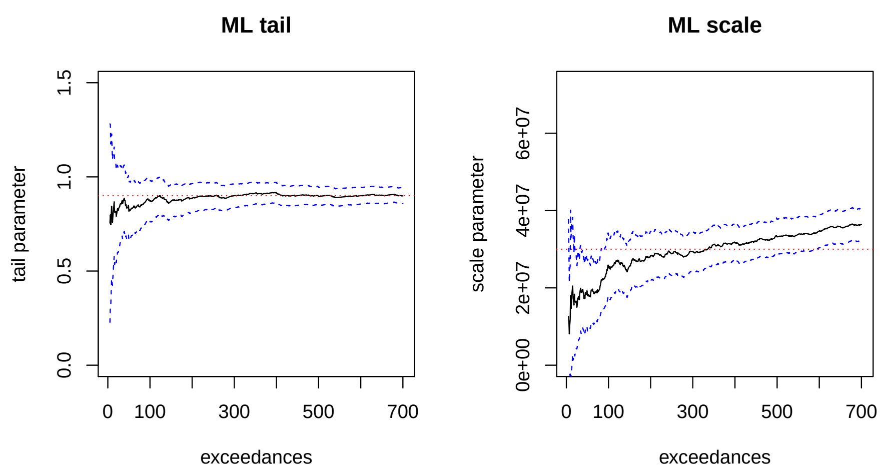

Figure 11: Stability plots for the tail and scale parameter of the Mittag-Leffler distribution of the Solar Flare dataset. Dotted horizontal lines are at $\beta = 0.85$ and $\sigma_0 = 3 \times 10^7$ seconds ≈ 0.95 years.

The bursty CTRE approach to model "non-Poissonian" threshold crossing times should be contrasted with the (now standard) approach of clusters of extremes, see e.g. Ferro and Segers (2003). In this approach, i.i.d. event sequences of magnitudes are generalized to stationary sequences of event magnitudes (subject to a mixing condition). The two approaches are fundamentally different: A clustering model assumes that each event belongs to one particular (latent) group of events. For bursts, however, the aim is to identify an underlying scale-free pattern in the event dynamics, which is often characteristic of complex systems. It is an interesting open problem to develop quality criteria, based e.g. on measures of surprise (Lee, Fan, and Sisson 2015), which guide an applied statistician in the choice between a clustering and a CTRE approach for a particular problem. Moreover, we believe it may be possible to unify the two approaches by considering CTREs based on MRPs with a *stationary*, rather than i.i.d., sequence of magnitudes.

Finally, a purely scale-free pattern for event times may be too rigid as an assumption for some bursty time series, because often the heavy-tailed character of the inter-arrival time distribution does not hold at all time scales; rather, it applies at short and intermediate time scales, and is truncated (or tempered, reverting to an exponential distribution) at very long time scales (see e.g. Mark M. Meerschaert, Roy, and Shao 2012; and Aban, Meerschaert, and Panorska 2006). In such situations, a "tempered" Mittag-Leffler distribution may provide a more realistic fit, which we aim to introduce in follow-up work.

Acknowledgements

The authors would like to thank Prof. Peter Scheffler for insights on stochastic process limits for CTRMs, Roloand Fried for discussion according the statistical methods and Gurtek Gill who helped create the MittagLeffleR R-package.

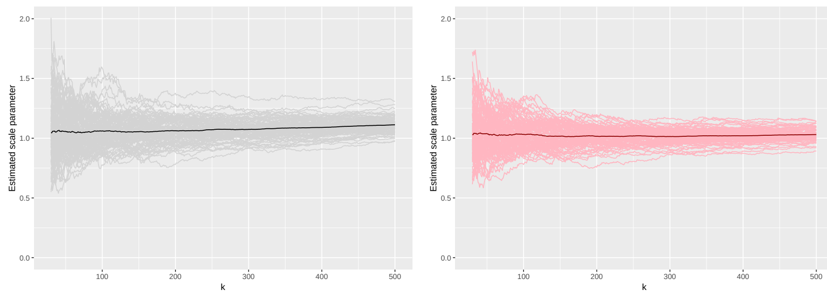

Figure 12: Stability Plots for the scale parameter for Example 1 (stable distributed waiting times, true scale parameter = 1).

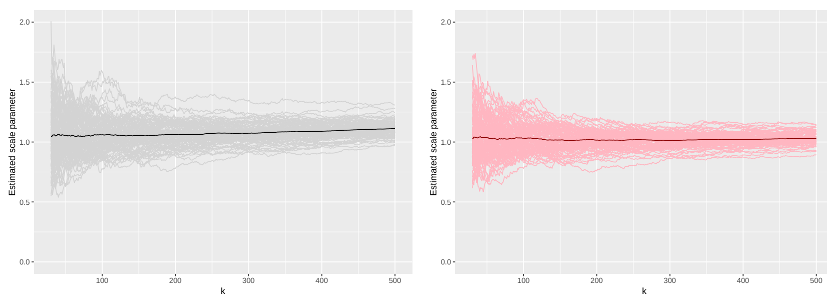

Figure 13: Stability Plots for the scale parameter m=100 simulation runs for Example 2 (stable distributed waiting times, true scale parameter = 1).

Appendix

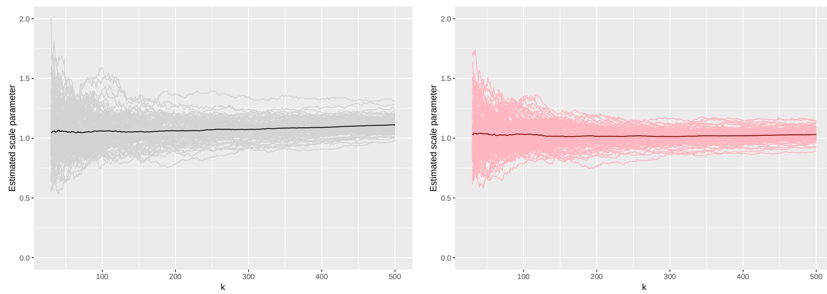

Figure 14: Stability Plots for the scale parameter for m=100 simulation runs for Example 3 (stable distributed waiting times, true scale parameter = 1).

References

Aban, Inmaculada B, Mark M Meerschaert, and Anna K Panorska. 2006. "Parameter Estimation for the Truncated Pareto Distribution." *J. Am. Stat. Assoc.* 101 (473): 270–77. https://doi.org/10.1198/016214505000000411.

Anderson, Kevin K. 1987. "Limit Theorems for General Shock Models with Infinite Mean Intershock Times." *J. Appl. Probab.* 24 (2): 449–56. http://www.jstor.org/stable/3214268.

Bagrow, James P., and Dirk Brockmann. 2013. "Natural emergence of clusters and bursts in network evolution." *Phys. Rev. X* 3 (2): 1–6. https://doi.org/10.1103/PhysRevX.3.021016.

Barabási, Albert László. 2005. "The origin of bursts and heavy tails in human dynamics." *Nature* 435 (May): 207–11. https://doi.org/10.1038/nature03459.

Basrak, Bojan, and Drago Špoljarić. 2015. "Extremes of random variables observed in renewal times." *Stat. Probab. Lett.* 97. Elsevier B.V.: 216–21. https://doi.org/10.1016/j.spl.2014.11.025.

Beirlant, Jan, Yuri Goegebeur, Johan Segers, and Jozef Teugels. 2006. *Statistics of extremes: theory and applications.* John Wiley & Sons.

Benson, David A, Rina Schumer, and Mark M Meerschaert. 2007. "Recurrence of extreme events with power-law interarrival times." *Geophys. Res. Lett.* 34 (l16404): DOI:10.1029/2007GL030767. https://doi.org/10.1029/2007GL030767.

Cahoy, Dexter O. 2013. "Estimation of Mittag-Leffler Parameters." *Commun. Stat. - Simul. Comput.* 42 (2): 303–15. https://doi.org/10.1080/03610918.2011.640094.

Cahoy, Dexter O, Vladimir V. Uchaikin, and W. A. Woyczynski Wojbor. 2010. "Parameter estimation for fractional Poisson processes." *J. Stat. Plan. Inference* 140 (11). Elsevier: 3106–20. https://doi.org/10.1016/j.jspi.2010.04.016.

Coles, S. 2001. *An Introduction to Statistical Modelling of Extreme Values.* London: Springer-Verlag.

Dennis, Brian R, L E Orwig, G S Kennard, G J Labow, R A Schwartz, A R Shaver, and A K Tolbert. 1991. "The complete Hard X Ray Burst Spectrometer event list, 1980-1989." https://umbra.nascom.nasa.gov/smm/hxrbs.html.

Esary, J. D., and A. W. Marshall. 1973. "Shock Models and Wear Processes." https://doi.org/10.1214/aop/1176996891.

Ferro, Christopher AT, and Johan Segers. 2003. "Inference for Clusters of Extreme Values." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65 (2). Wiley Online Library: 545–56.

Gill, Gurtek, and Peter Straka. 2017. *MittagLeffleR: Using the Mittag-Leffler Distributions in R.* https://strakaps.github.io/MittagLeffleR/.

Gut, Allan, and Jürg Hüsler. 1999. "Extreme Shock Models." *Extremes*, no. 1983: 295–307. http://link.springer.com/article/10.1023/A:1009959004020.

Haubold, H.J., A. M. Mathai, and R. K. Saxena. 2011. "Mittag-Leffler Functions and Their Applications." *J. Appl. Math.* 2011: 1–51. https://doi.org/10.1155/2011/298628.

Hawkes, Alan G. 1971. "Point spectra of some mutually exciting point processes." *J. R. Stat. Soc. Ser. B Stat. Methodol.* JSTOR, 438–43.

Hees, Katharina, and Hans-Peter Scheffler. 2016. "On joint sum/max stability and sum/max domains of attraction," June, 1–31. http://arxiv.org/abs/1606.03109.

———. 2017. "Coupled Continuous Time Random Maxima." *Extremes*, no. June (September). Extremes: 1–24. https://doi.org/10.1007/s10687-017-0304-6.

Hsing, T, J Hüsler, and M R Leadbetter. 1988. "On the Exceedance Point Process for a Stationary Sequence" 78: 97–112.

Karsai, Márton, Kimmo Kaski, Albert László Barabási, and János Kertész. 2012. "Universal features of correlated bursty behaviour." *Sci. Rep.* 2. https://doi.org/10.1038/srep00397.

Karsai, Márton, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, Albert László Barabási, and J. Saramäki. 2011. "Small but slow world: How network topology and burstiness slow down spreading." *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* 83: 1–4. https://doi.org/10.1103/PhysRevE.83.025102.

Kozubowski, Tomasz J. 2001. "Fractional Moment Estimation of Linnik and Mittag-Leffler Parameters." *Mathematical and Computer Modelling* 34 (9-11). Elsevier: 1023–35.

Laskin, Nick. 2003. "Fractional Poisson process." *Commun. Nonlinear Sci. Numer. Simul.* 8 (3-4): 201–13. https://doi.org/10.1016/S1007-5704(03)00037-6.

Lee, J., Y. Fan, and S. A. Sisson. 2015. "Bayesian threshold selection for extremal models using measures of surprise." *Comput. Stat. Data Anal.* 85. Elsevier B.V.: 84–99. https://doi.org/10.1016/j.csda.2014.12.004.

Meerschaert, Mark M, Erkan Nane, and P. Vellaisamy. 2011. "The fractional Poisson process and the inverse stable subordinator." *Electron. J. Probab.* 16: 1600–1620. https://doi.org/10.1214/EJP.v16-920.

Meerschaert, Mark M., Parthanil Roy, and Qin Shao. 2012. "Parameter estimation for exponentially tempered power law distributions." *Commun. Stat. - Theory Methods* 41 (10): 1839–56. https://doi.org/10.1080/03610926.2011.552828.

Meerschaert, Mark M, and Alla Sikorskii. 2012. *Stochastic Models for Fractional Calculus.* Vol. 43. Walter de Gruyter.

Meerschaert, Mark M, and Stilian A Stoev. 2008. "Extremal limit theorems for observations separated by random power law waiting times." *J. Stat. Plan. Inference* 139 (7): 2175–88. https://doi.org/10.1016/j.jspi.2008.10.005.

Min, Byungjoon, K. I. Goh, and Alexei Vazquez. 2011. "Spreading dynamics following bursty human activity patterns." *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* 83 (3): 2–5. https://doi.org/10.1103/PhysRevE.83.036102.

Oliveira, J, and Albert László Barabási. 2005. "Darwin and Einstein correspondence patterns." *Nature* 437 (October): 1251. https://doi.org/10.0138/4371251a.

Omi, Takahiro, and Shigeru Shinomoto. 2011. "Optimizing Time Histograms for Non-Poissonian Spike Trains." *Neural Comput.* 23 (12): 3125–44. https://doi.org/10.1162/NECO_a_00213.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Resnick, Sidney I. 1997. "Heavy tail modeling and teletraffic data." *Ann. Stat.* 25 (5): 1805–49. https://doi.org/10.1214/aos/1069362376.

Samorodnitsky, Gennady, and Murad S Taqqu. 1994. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Stochastic Modeling. London: Chapman Hall.

Shanthikumar, J. George, and Ushio Sumita. 1983. "General shock models associated with correlated renewal sequences." *J. Appl. Probab.* 20 (3): 600–614. http://www.jstor.org/stable/3213896.

———. 1984. "Distribution Properties of the System Failure Time in a General Shock Model." *Adv. Appl. Probab.* 16 (2): 363–77. http://www.jstor.org/stable/1427074.

———. 1985. "A class of correlated cumulative shock models." *Adv. Appl. Probab.* 17 (2): 347–66. http://www.jstor.org/stable/1427145.

Silvestrov, Dmitrii S. 2002. *Limit Theorems for Randomly Stopped Stochastic Processes*. Springer (Berlin, Heidelberg).

Silvestrov, Dmitrii S, and Jozef L. Teugels. 2004. "Limit theorems for mixed max-sum processes with renewal stopping." *Ann. Appl. Probab.* 14 (4): 1838–68. https://doi.org/10.1214/105051604000000215.

Vajna, Szabolcs, Bálint Tóth, and János Kertész. 2013. "Modelling bursty time series." *New J. Phys.* 15 (10): 103023. https://doi.org/10.1088/1367-2630/15/10/103023.

Vasquez, a, J G Oliveira, Z Dezso, K-I Goh, I Kondor, and Albert László Barabási. 2006. "Modeling bursts and heavy tails in human dynamics." *Phys. Rev. E* 73: 361271–3612718.

Vazquez, Alexei, Balázs Rácz, András Lukács, and Albert László Barabási. 2007. "Impact of non-poissonian activity patterns on spreading processes." *Phys. Rev. Lett.* 98 (APRIL): 1–4. https://doi.org/10.1103/PhysRevLett.98.158702.

Wuertz, Diethelm, Martin Maechler, and Rmetrics core team members. 2016. *Stabledist: Stable Distribution Functions*. https://CRAN.R-project.org/package=stabledist.