CTRM Statistics

October 7, 2016

Abstract

We fit a max-renewal process, (aka 'CTRM', Continuous Time Random Maxima) to data.

1 Introduction

Time series displaying inhomogeneous behaviour have received strong interest in the recent statistical physics literature, [Bar05, OB05, VOD+06, VRLB07, OS11, MGV11, KKP+11, BB13], and have been observed in the context of earthquakes, sunspots, neuronal activity, human communication etc., see [KKBK12, VTK13] 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 processes [Haw71]), there seems to be one salient feature which very typically indicates the departure from temporal homogeneity: A heavy-tailed distribution of waiting times [VOD+06, KKBK12, VTK13]. A simple renewal process with heavy-tailed waiting times captures these dynamics. For many systems, the renewal property is appropriate, as can be checked by a simple test: the dynamics do not change significantly if the waiting times are randomly reshuffled [KKBK12].

When a magnitude can be assigned to each event in the renewal process, such as e.g. for earthquakes, sun flares, neuron voltages or the impact of an email, two natural and basic questions to ask are: What is the distribution of the largest event up to a given time t? What is the probability that an event exceeds a given level ℓ within the next t units of time? A probabilistic extreme value model which assumes that the events form a renewal process is available in the literature. This model has been studied under the names "Continuous Time Random Maxima process" (CTRM) [BSM07, **?**, HS15, **?**], "Max-Renewal process" [Sil02, ST04, BŠ14], and "Shock process" [EM73, SS83, SS84, SS85, And87, Gut99]. This article aims to develop concrete statistical inference methods for this model, a problem which has seemingly received little attention by the statistical community.

2 CTRM processes

The Continuous Time Random *Walk* (CTRW) has been a highly successful model for anomalous diffusion in the past two decades [MK00, HTSL10], likely due to its tractable and flexible scaling properties. The stochastic process we study in this article is conceptually very close to the CTRW, since essentially the jumps J_k are reinterpreted as magnitudes, and instead of the cumulative sum, one tracks the cumulative maximum. Similarly tractable scaling properties apply to the CTRM, see below. For the above reasons, we use the name CTRM in this article.

Definition 2.1. *Assume i.i.d. pairs of random variables* (J_k, W_k)*,* $k = 1, 2, \dots$ *where* $W_k > 0$ *represents the inter-arrival times of certain events and* $J_k \in \mathbb{R}$ *the corresponding event magnitudes. Write*

$$N(t) = \max\{n \in \mathbb{N} : W_1 + \ldots + W_n \le t\} \tag{2.1}$$

for the renewal process associated with the W_k *(where the maximum of the empty set is set to* 0*). Then the process*

$$M(t) = \bigvee_{k=1}^{N(t)} J_k = \max\{J_k : k = 1, \dots, N(t)\}, \quad t \ge 0. \tag{2.2}$$

is called a CTRM (Continuous Time Random Maxima process).

It is clear that the sample paths of $N(t)$ and $M(t)$ are right-continuous with left-hand limits. If W_k is interpreted as the time *leading up* to the event with magnitude J_k, then $M(t)$ is indeed the largest magnitude up to time t. The alternative case where W_k represents the inter-arrival time *following* J_k is termed "second type" (in the shock model literature) or OCTRM (overshooting CTRM), and the largest magnitude up to time t is then given by

$$\tilde{M}(t) = \bigvee_{k=1}^{N(t)+1} J_k, \quad t \ge 0. \tag{2.3}$$

Finally, the model is called *coupled* when W_k and J_k are not independent. In this article we focus on the uncoupled case, for which it can be shown that the processes $M(t)$ and $\tilde{M}(t)$ have the same limiting distributions at large times [**?**], and hence we focus on the CTRM $M(t)$.

We aim to make inference on the distribution of the following quantities:

Definition 2.2. *Let* $M(t)$ *be an uncoupled CTRM whose magnitudes* J_k *are supported on the interval* $[x_0, x_F]$*. Then the exceedance time of level* $\ell \in [x_0, x_F]$ *is the random variable*

$$T_\ell = \inf\{t : M(t) > \ell\}$$

and the exceedance is

$$X_\ell = M(T_\ell) - \ell.$$

Lemma 2.3. *Given a level $\ell \in [x_0, x_F]$, exceedance X_ℓ and exceedance time T_ℓ are independent. Moreover,*

$$\mathbf{P}[X_\ell > x] = \frac{\overline{F}_J(\ell + x)}{\overline{F}_J(\ell)}, \quad x > 0,$$
$$\mathbf{P}[T_\ell > t] = \overline{F}_J(\ell) \sum_{n=1}^{\infty} \int_0^\infty \overline{F}_W(t - t')\mathbf{P}[L_{n-1} \le \ell]\mathbf{P}[S_{n-1} \in dt'], \quad t > 0$$

where $L_n = \bigvee_{k=1}^n J_k$, $L_0 = x_0$, and $S_n = \sum_{k=1}^n W_k$, $S_0 = 0$.

Proof. Let $\tau_\ell = \min\{k : J_k > x\}$. Then

$$\begin{aligned} T_\ell > t, X_\ell > x \iff S_{\tau_\ell} > t, L_{\tau_\ell} > \ell + x \iff \exists n : L_{n-1} \le \ell, J_n > \ell + x, S_n > t \\ \iff \exists n : L_{n-1} \le \ell, J_n > \ell + x, W_n > t - S_{n-1} \end{aligned}$$

and such n must be unique, since $x > 0$. Thus

$$\begin{aligned} &\mathbf{P}[T_\ell > t, X_\ell > x] = \sum_{n=1}^{\infty} \mathbf{P}[J_n > \ell + x, W_n > t - S_{n-1}, L_{n-1} \le \ell] \\ &= \sum_{n=1}^{\infty} \int_0^\ell \int_0^\infty \mathbf{P}[J_n > \ell + x, W_n > t - t' | L_{n-1} = m', S_{n-1} = t']\mathbf{P}[L_{n-1} \in dm', S_{n-1} \in dt'] \\ &= \sum_{n=1}^{\infty} \int_0^\infty \overline{F}_J(\ell + x)\overline{F}_W(t - t')\mathbf{P}[L_{n-1} \le \ell]\mathbf{P}[S_{n-1} \in dt'] \end{aligned}$$

since the sequences J_k and W_k are i.i.d. and independent of each other. Letting $t \downarrow 0$, we see

$$\begin{aligned} \mathbf{P}[X_\ell > x] &= \sum_{n=1}^{\infty} \overline{F}_J(\ell + x)\mathbf{P}[L_{n-1} \le \ell] = \overline{F}_J(\ell + x) \sum_{n=1}^{\infty} F_J(\ell)^{n-1} \\ &= \frac{\overline{F}_J(\ell + x)}{\overline{F}_J(\ell)}, \end{aligned}$$

and letting $x \downarrow 0$, one gets $\mathbf{P}[T_\ell > t]$. One checks that $\mathbf{P}[T_\ell > t, X_\ell > x] = \mathbf{P}[T_\ell > t]\mathbf{P}[X_\ell > x]$, implying independence. $\square$

The distribution of the exceedance is hence, as expected, simply the conditional distribution of a magnitude J_k given $J_k > \ell$, and may hence be modeled by a generalized Pareto (GP) distribution [Col01]. The exceedance times, and hence the temporal aspect of the dynamics, are independent of the exceedances, and may be looked at separately.

3 Fractional Time Scale

As mentioned in the introduction, we assume a heavy-tailed distribution for the waiting times, more precisely that the tail function $\overline{F}_W := 1 - F_W$ of the CDF of W is regularly varying:

$$t \mapsto \overline{F}_W(t) \in RV(-\beta), \quad \beta \in (0,1) \tag{3.1}$$

meaning that [Sen76, MS01]

$$\lim_{t\to\infty} \frac{\overline{F}_W(\lambda t)}{\overline{F}_W(t)} = \lambda^{-\beta}.$$

By [MS01, Cor. 8.2.19] (also compare [Whi01, Th. 4.5.1]), this is equivalent to W_1 being in the (sum-) domain of attraction of a positively skewed stable law D with stability parameter β, which means the weak convergence

$$b(n)(W_1 + \ldots + W_n) \Rightarrow D, \quad n \to \infty. \tag{3.2}$$

The i.i.d. nature of the W_k allows for a more general, functional limit theorem [MS01, Ex. 11.2.18], [Whi01, Th. 4.5.3]:

$$b(c)(W_1 + \ldots + W_{\lfloor cr \rfloor}) \Rightarrow D(cr), \quad c \to \infty, \tag{3.3}$$

where $\{D(r)\}_{t\geq 0}$ is a stable subordinator, defined via

$$\mathbf{E}[e^{-\lambda D(r)}] = e^{-r\lambda^\beta}, \tag{3.4}$$

where $\lfloor r \rfloor$ denotes the largest integer not larger than r, and where convergence holds on the stochastic process level, i.e. in the sense of weak convergence in Skorokhod space endowed with the J_1 topology [Whi01, Sec. 3.3].

The renewal process paths $N(t)$ result, up to a constant, from inverting the non-decreasing sample paths of the sum $r \mapsto W_1 + \ldots + W_{\lfloor r \rfloor}$. Accordingly, the scaling limit of the renewal process is [MS04]

$$N(ct)/\tilde{b}(c) \Rightarrow E(t), \quad c \to \infty \tag{3.5}$$

where $E(t)$ denotes the inverse stable subordinator [MS13]

$$E(t) = \inf\{r : D(r) > t\}, \quad t \geq 0, \tag{3.6}$$

and where $\tilde{b}(c)$ is asymptotically inverse to $1/b(c)$, in the sense of [Sen76, p.20]:

$$1/b(\tilde{b}(c)) \sim c \sim \tilde{b}(1/b(c)) \tag{3.7}$$

where a $\sim$ symbol indicates that the quotient of both sides converges to 1 as $c \to \infty$. Note that $\tilde{b}(c) \in RV(\beta)$.

The inverse stable subordinator $E(t)$ governs the temporal dynamics of the scaling limit of the CTRM process $M(t)$, see Theorem **??**. It is self-similar with exponent β [MS04], non-decreasing, and the (regenerative, random) set $\mathcal{R}$ of its points of increase is a fractal with dimension β [Ber99]. $E(t)$ is a model for time series with intermittent, 'bursty' behaviour, for the following reasons:

i) Conditional on $t \geq 0$ being a point of increase, any interval $(t, t+\epsilon)$ almost surely contains uncountably many other points of $\mathcal{R}$;

ii) $\mathcal{R}$ has Lebesgue measure 0, and hence $E(t)$ is constant at any "randomly" chosen time t.

Having only two parameters ($\beta \in (0,1)$ and a scale parameter) the inverse stable subordinator hence models scaling limits of heavy-tailed waiting times parsimoniously.

4 Generalized Extreme Value Distributions

We turn our attention to the exceedances X_ℓ of a given high threshold ℓ 'close' to the right end-point x_F of F_J. Any non-degenerate distribution on $\mathbb{R}$

5 CTRM limit

The timing of maxima. The distribution of the largest magnitudes is well approximated by a generalized extreme value (GEV) distribution F if the number of events is large [Col01]. Consider again the light-tailed inter-arrival time case: Exceedances of large levels ℓ with small occurrence probability $p = 1 - F(\ell)$ occur, on average every $1/p$ observations, and the average return period is quickly seen to be $\mathbf{E}[W_k]/p$. This provides an intuitive interpretation of magnitudes, e.g. "a flood exceeding level ℓ occurs only every 100 years on average." Moreover, the lightness of the tails of W_k implies, as mentioned above, that events occur in a Poisson process fashion. The memory of large events in the past is negligible for the prediction of large events in the future. For heavy-tailed inter-arrival times, however, the Markov property does not hold, and time averages of return periods are not defined, since their moments are infinite.

Aims. The goal of our statistical method is to make inferences about the dynamics of $M(t)$, given data (W_k, J_k). Since the waiting times J_k are not assumed exponential, $M(t)$ is not Markov, and the same applies to $A(E(t))$.

As was shown by [And87] (see also [MS08]), exceedance times are Mittag-Leffler distributed, and thus form a *fractional* Poisson process [Las03]. Due to recent progress in the estimation of Mittag-Leffler distributions [Cah13, CUW10] we are able to model the recurrence periods, and thus to give predictive distributions for future exceedance times, using the semi-Markov property of fractional Poisson processes.

[**?**] have shown that for large times t, the process $M(t)$ converges to a (rescaled) time-changed extreme-value process $A(E(t))$. Here, $A(t)$ is an extreme-value process in the sense of [Lam64] and [Res13], and

Any probability distribution with cumulative distribution function (CDF) F on $\mathbb{R}$ lies in the max-domain of attraction of a generalized extreme value (GEV) distribution with CDF Λ [BGST06]. That is, there exist functions $a : \mathbb{R}^+ \to \mathbb{R}^+$ and $d : \mathbb{R}^+ \to \mathbb{R}^+$ such that

$$F(a(n)x + b(n))^n \to \Lambda(x), \quad n \to \infty, \quad x \in (x_L, x_R), \tag{5.1}$$

where x_L and x_R denote the left and right endpoint of the GEV distribution.

6 Statistical Method

7 Simulated Data

8 Real Data

9 Conclusion

References

[And87] Kevin K Anderson. Limit Theorems for General Shock Models with Infinite Mean Intershock Times. *J. Appl. Probab.*, 24(2):449–456, 1987.

[Bar05] Albert László Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(May):207–211, 2005.

[BB13] James P. Bagrow and Dirk Brockmann. Natural emergence of clusters and bursts in network evolution. *Phys. Rev. X*, 3(2):1–6, 2013.

[Ber99] Jean Bertoin. *Subordinators: examples and applications*, volume 1717 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

[BGST06] Jan Beirlant, Yuri Goegebeur, Johan Segers, and Jozef Teugels. *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.

[BŠ14] Bojan Basrak and Drago Špoljarić. Extremal behaviour of random variables observed in renewal times. jun 2014.

[BSM07] David A Benson, Rina Schumer, and Mark M Meerschaert. Recurrence of extreme events with power-law interarrival times. *Geophys. Res. Lett.*, 34(l16404):DOI:10.1029/2007GL030767, aug 2007.

[Cah13] Dexter O. Cahoy. Estimation of Mittag-Leffler Parameters. *Commun. Stat. - Simul. Comput.*, 42(2):303–315, feb 2013.

[Col01] S Coles. *An Introduction to Statistical Modelling of Extreme Values*. Springer-Verlag, London, 2001.

[CUW10] Dexter O. Cahoy, Vladimir V. Uchaikin, and W. A. Woyczynski Wojbor. Parameter estimation for fractional Poisson processes. *J. Stat. Plan. Inference*, 140(11):3106–3120, 2010.

[EM73] J. D. Esary and A. W. Marshall. Shock Models and Wear Processes, 1973.

[Gut99] Allan Gut. Extreme Shock Models. *Extremes*, (1983):295–307, 1999.

[Haw71] Alan G Hawkes. Point spectra of some mutually exciting point processes. *J. R. Stat. Soc. Ser. B*, pages 438–443, 1971.

[HS15] Katharina Hees and H.P. Scheffler. Coupled continuous time random maxima. pages 1–24, 2015.

[HTSL10] B.I. Henry, T. A.M. Langlands, Peter Straka, and T.A.M. Langlands. An introduction to fractional diffusion. In R L. Dewar and F Detering, editors, *Complex Phys. Biophys. Econophysical Syst. World Sci. Lect. Notes Complex Syst.*, volume 9 of *World Scientific Lecture Notes in Complex Systems*, pages 37–90, Singapore, 2010. World Scientific.

[KKBK12] Márton Karsai, Kimmo Kaski, Albert László Barabási, and János Kertész. Universal features of correlated bursty behaviour. *Sci. Rep.*, 2, 2012.

[KKP$^+$11] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, Albert László Barabási, and J. Saramäki. Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, 83:1–4, 2011.

[Lam64] J Lamperti. On extreme order statistics. *Ann. Math. Stat.*, 35(4):1726–1737, 1964.

[Las03] Nick Laskin. Fractional Poisson process. *Commun. Nonlinear Sci. Numer. Simul.*, 8(3-4):201–213, 2003.

[MGV11] Byungjoon Min, K. I. Goh, and Alexei Vazquez. Spreading dynamics following bursty human activity patterns. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, 83(3):2–5, 2011.

[MK00] Ralf Metzler and Joseph Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.*, 339(1):1–77, dec 2000.

[MS01] Mark M Meerschaert and H.P. Scheffler. *Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice.* Wiley-Interscience, New York, first edition, jul 2001.

[MS04] Mark M Meerschaert and H.P. Scheffler. Limit Theorems for Continuous-Time Random Walks with Infinite Mean Waiting Times. *J. Appl. Probab.*, 41(3):623–638, sep 2004.

[MS08] Mark M Meerschaert and Stilian A Stoev. Extremal limit theorems for observations separated by random power law waiting times. *J. Stat. Plan. Inference*, 139(7):2175–2188, jul 2008.

[MS13] Mark M Meerschaert and Peter Straka. Inverse Stable Subordinators. *Math. Model. Nat. Phenom.*, 8(2):1–16, apr 2013.

[OB05] J Oliveira and Albert László Barabási. Darwin and Einstein correspondence patterns. *Nature*, 437(October):1251, 2005.

[OS11] Takahiro Omi and Shigeru Shinomoto. Optimizing Time Histograms for Non-Poissonian Spike Trains. *Neural Comput.*, 23(12):3125–3144, 2011.

[Res13] Sidney I Resnick. *Extreme values, regular variation and point processes.* Springer, 2013.

[Sen76] E Seneta. *Regularly Varying Functions*, volume 508 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 1976.

[Sil02] Dmitrii S Silvestrov. *Limit Theorems for Randomly Stopped Stochastic Processes.* Springer (Berlin, Heidelberg), 2002.

[SS83] J. George Shanthikumar and Ushio Sumita. General shock models associated with correlated renewal sequences. *J. Appl. Probab.*, 20(3):600–614, 1983.

[SS84] J. George Shanthikumar and Ushio Sumita. Distribution Properties of the System Failure Time in a General Shock Model. *Adv. Appl. Probab.*, 16(2):363–377, 1984.

[SS85] J. George Shanthikumar and Ushio Sumita. A class of correlated cumulative shock models. *Adv. Appl. Probab.*, 17(2):347–366, 1985.

[ST04] Dmitrii S Silvestrov and Jozef L. Teugels. Limit theorems for mixed max-sum processes with renewal stopping. *Ann. Appl. Probab.*, 14(4):1838–1868, nov 2004.

[VOD+06] a Vasquez, J G Oliveira, Z Dezso, K-I Goh, I Kondor, and Albert László Barabási. Modeling bursts and heavy tails in human dynamics. *Phys. Rev. E*, 73:361271–3612718, 2006.

[VRLB07] Alexei Vazquez, Balázs Rácz, András Lukács, and Albert László Barabási. Impact of non-poissonian activity patterns on spreading processes. *Phys. Rev. Lett.*, 98(APRIL):1–4, 2007.

[VTK13] Szabolcs Vajna, Bálint Tóth, and János Kertész. Modelling bursty time series. *New J. Phys.*, 15(10):103023, oct 2013.

[Whi01] Ward Whitt. *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and their Application to Queues.* Springer, New York, 1st edition, nov 2001.