STATISTICS OF BURSTY EVENTS

Smarak Nayak

Supervisor: Dr. Peter Straka

School of Mathematics and Statistics
UNSW Australia

October 2016

Submitted in partial fulfillment of the requirements of the degree of
Bachelor of Science with Honours

Plagiarism statement

I declare that this thesis is my own work, except where acknowledged, and has not been submitted for academic credit elsewhere.

I acknowledge that the assessor of this thesis may, for the purpose of assessing it:

- Reproduce it and provide a copy to another member of the University; and/or,
- Communicate a copy of it to a plagiarism checking service (which may then retain a copy of it on its database for the purpose of future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct, and am aware of any potential plagiarism penalties which may apply.

By signing this declaration I am agreeing to the statements and conditions above.

Signed: ____________________________ Date: ________________

Acknowledgements

By far the greatest thanks must go to my supervisor for the guidance, care and support they provided.

Thanks must also go to Emily, Michelle, John and Alex who helped by proof-reading the document in the final stages of preparation.

Although I have not lived with them for a number of years, my family also deserve many thanks for their encouragement.

Thanks go to Robert Taggart for allowing his thesis style to be shamelessly copied.

Fred Flintstone, 2 November 2015.

Abstract

The prediction of extreme events resulting from human behaviour is becoming significant in areas as far reaching as the control of disease spread, resource allocation, and emergency response. Classical methods in extreme value theory fail to recognise that most human created events do not occur uniformly, but in bursts. In this thesis, we aim to design statistical methods for the inference and prediction of Continuous Time Random Maxima (CTRM).

Contents

Chapter 1 Introduction 1

Chapter 2 Preliminaries 2

2.1 Probability and Measure Theory 2

2.2 Stochastic Processes 4

2.3 Weak Convergence 5

2.4 The Skorokhod Space 6

2.5 Central Limit Theorems 8

Chapter 3 Limit Theorems 10

3.1 Problem Formulation 10

3.2 Brownian Motion as a Stochastic Process Limit 13

3.3 Stable Distributions 16

3.4 Scaling Limit of the Sum of Waiting Times 19

3.5 Scaling Limit of the Maximum of Event Magnitudes 23

3.6 Scaling Limit of CTRM 27

Chapter 4 Distribution of CTRM 29

4.1 Distribution of Exceedances 29

4.2 Distribution of Exceedance Durations 29

Chapter 5 Statistical Inference 30

5.1 Log Moment Estimator for Mittag-Leffler Random Variables 30

5.2 Maximum Likelihood Estimator for Generalised Pareto Random Variables 30

5.3 Stability of Parameter Estimates 30

5.4 Goodness of Fit 30

Chapter 6 Model Fitting 31

6.1 Model Fit of Simulated Data 31

6.2 Model Fit of Real Data 33

Chapter 7 Conclusion 34

References 35

Chapter 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 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.

Chapter 2

Preliminaries

In this chapter, we introduce the notation and review the background theory of probability and stochastic processes that are necessary in developing the statistical methods outlined in this thesis.

2.1 Probability and Measure Theory

Definition 2.1.1. *Let Ω be a nonempty set and let $\mathcal{F}$ be a collection of subsets of Ω. We say that $\mathcal{F}$ is a **σ-algebra** if it satisfies the following conditions*

1. *The empty set $\emptyset \in \mathcal{F}$,*
2. *If $A \in \mathcal{F}$ then the complement $A^c \in \mathcal{F}$, and*
3. *If $A_1, A_2, \ldots \in \mathcal{F}$ then their union $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.*

It follows that a σ-algebra is also closed under intersection.

Definition 2.1.2. *Let Ω be a nonempty set and let $\mathcal{F}$ be a σ-algebra on Ω. Then the pair $(\Omega, \mathcal{F})$ is called a **measurable space**.*

Definition 2.1.3. *Let $(\Omega, \mathcal{F})$ be a measurable space. A **measure** μ is a real-valued set function on $\mathcal{F}$ satisfying the following conditions*

1. *For every set $A \in \mathcal{F}$, $\mu(A) \geq 0$, and*
2. *If $A_1, A_2, \ldots$ is a sequence of disjoint sets in $\mathcal{F}$ then*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Now that the concept of a measure has been defined we can interpret a σ-algebra as the set of all events that can be measured from the underlying set Ω.

Definition 2.1.4. *Let* $(\Omega, \mathcal{F})$ *be a measurable space. A* ***probability measure*** $\mathbb{P}$ *is a measure such that* $\mathbb{P}(\Omega) = 1$.

Definition 2.1.5. *Let* μ *be a measure defined on the measurable space* $(\Omega, \mathcal{F})$*. We call the triplet* $(\Omega, \mathcal{F}, \mu)$ *a* ***measure space****. In particular, if* $\mathbb{P}$ *is a probability measure then we call* $(\Omega, \mathcal{F}, \mathbb{P})$ *a* ***probability space****.*

A particular example of a probability measure that we will use later on is the Dirac measure.

Definition 2.1.6. *Let* $(\Omega, \mathcal{F})$ *be some measurable space. A* ***Dirac measure*** δ_x *is a measure defined on a given point* $x \in \Omega$ *and a set* $A \in \mathcal{F}$

$$\delta_x(A) = \begin{cases} 0, & x \notin A; \\ 1, & x \in A. \end{cases}$$

A useful result of Dirac measures is the identity

$$\int f(y)\delta_x(dy) = f(x).$$

The Dirac measure extends the concept of indicator functions to sets. Another example of a measure is the Lebesgue measure, which is the standard measure on the Euclidean space $\mathbb{R}^d$. The rigorous definition of a Lebesgue measure is fairly analytical and is omitted. The Lebesgue measure is essentially an extension of the standard notions of length, area and volume to d-dimensional Euclidean spaces.

Definition 2.1.7. *Let* $(\Omega, \mathcal{F})$ *and* $(E, \mathcal{E})$ *be measurable spaces. Then a function* $f : \Omega \to E$ *is* $(\mathcal{F}, \mathcal{E})$*-**measurable** if* $f^{-1}(S) \in \mathcal{F}$ *for every* $S \in \mathcal{E}$.

Definition 2.1.8. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* $(E, \mathcal{E})$ *a measurable space. A* ***random variable*** *is a function* $X : \Omega \to E$ *that is* $(\mathcal{F}, \mathcal{E})$*-measurable. We call* $(E, \mathcal{E})$ *the* ***state space****.*

Definition 2.1.9. *Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Its* ***characteristic function*** $\phi_X(s) : \Omega \to \mathbb{C}$ *is defined by*

$$\begin{aligned}\phi_X(s) &= \mathbb{E}[e^{isX}] \\ &= \int_\Omega e^{isX(\omega)}\mathbb{P}(d\omega)\end{aligned}$$

for every $s \in \mathbb{R}$.

2.2 Stochastic Processes

Generally speaking, a stochastic process $\{X_t\}_{t\geq 0}$ is a collection of random variables that represent the evolution of some object over time. As this thesis focuses heavily on stochastic processes it is important that we rigorously define what a stochastic process is.

Definition 2.2.1. *A* ***stochastic process*** *is a collection of random variables $\{X_t\}_{t\in\mathcal{I}}$ that is defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and is indexed by an ordered set $\mathcal{I}$.*

If we fix a $t_0 \in \mathcal{I}$ then $X(t_0, \omega)$ is simply a random variable. If we fix a $\omega_0 \in \Omega$ then $X(t, \omega_0)$ is a function with respect to t and is the trajectory of a single realisation of the stochastic process $X(t)$. From now on we will only consider stochastic processes where our ordered set $\mathcal{I} = \mathbb{R}^+$ represents an interval in continuous time. We now write $\{X(t)\}_{t\geq 0} : \mathbb{R}^+ \times \Omega \to \mathbb{R}$.

Definition 2.2.2. *A stochastic process $\{X(t)\}_{t\geq 0}$ has a* ***jump*** *at time t if the process is discontinuous at that point. More precisely $X(t) \neq X_-(t)$.*

Definition 2.2.3. *A stochastic process $\{X(t)\}_{t\geq 0}$ is* ***càdlàg*** *(derived from the French term "continue à droite, limite à gauche") if it is continuous from the right*

$$\lim_{\epsilon\to 0+} X(t+\epsilon) = X(t),$$

with left hand limits

$$\lim_{\epsilon\to 0+} X(t-\epsilon) = X(t-).$$

Definition 2.2.4. *A **filtration** $\{\mathcal{F}_t\}_{t\geq 0}$ is a collection of σ-algebras such that for any $0 \leq s \leq t$, $\mathcal{F}_s \subseteq \mathcal{F}_t$.*

Note that with each successive σ-algebra the amount of sets that can be measured is non-decreasing. Thus a filtration can be interpreted as the accumulation of information about our stochastic process with respect to time. We write $\sigma(X_s, s \leq t)$ as the smallest σ-algebra generated by $\{X_s : s \leq t\}$.

Definition 2.2.5. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with the filtration $\mathcal{F} = \{\mathcal{F}\}_{t\geq 0}$ and let $(E, \mathcal{E})$ a measurable space. We say the stochastic process $\{X_t\}_{t\geq 0}$ is **$\mathcal{F}$-adapted** if for every $t \geq 0$, $X_t : \Omega \to E$ is $(\mathcal{F}_t, \mathcal{E})$ - measurable.*

Definition 2.2.6. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A real valued stochastic process $\{W_t\}_{t\geq 0}$ with continuous paths is a **standard Brownian motion** if for all $0 < s < t$,*

- *$W_0 = 0$ a.s.,*
- *$W_t - W_s \sim N(0, t-s)$, and*
- *the increments $W_t - W_s$ are independent. That is, $W_t - W_s$ is independent of $\sigma(W_r, r \leq s)$.*

2.3 Weak Convergence

The notion of weak convergence or convergence in distribution is also required to understand results utilised in this thesis.

Definition 2.3.1. *A **metric** d on a set E is a non-negative real-valued function on the product space $E \times E$ such that*

- *$d(x, y) = 0$ iff $x = y$,*
- *$d(x, y) = d(y, x)$,*
- *$d(x, z) \leq d(x, y) + d(y, z)$.*

A metric is a function that defines the distance between two points in a set.

Definition 2.3.2. *Let E be a set and let d be a metric on E. Then the ordered pair (E,d) is a **metric space**.*

Definition 2.3.3. *A sequence $\{x_n\}_{n\geq 1}$ in a metric space (E,d)* ***converges to a limit*** *$x \in E$ if, for all $\epsilon > 0$, there exists an integer n_0 such that $d(x_n, x) < \epsilon$ for all $n \geq n_0$.*

Definition 2.3.4. *A sequence of probability measures $\{\mathbb{P}_n\}_{n\geq 1}$ on the measurable space $(\Omega, \mathcal{F})$* ***converges weakly*** *to a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ if*

$$\lim_{n\to\infty} \int_\Omega f d\mathbb{P}_n = \int_\Omega f d\mathbb{P}$$

for all continuous and bounded real-valued functions f on Ω. We then write

$$\mathbb{P}_n \to \mathbb{P}.$$

Note that this definition is just 2.3.3 with E as the set of all probability measures on the measurable space $(\Omega, \mathcal{F})$ and $d(\mathbb{P}, \mathbb{Q}) = |\int_\Omega f d\mathbb{P} - \int_\Omega f d\mathbb{Q}|$, where $\mathbb{Q}$ is another probability measure on $(\Omega, \mathcal{F})$.

Definition 2.3.5. *Let $X_n : \Omega \to E$ be a sequence of random variables with a corresponding sequence of probability measures $\{\mathbb{P}_{X_n}\}_{n\geq 1}$ on the measurable space $(\Omega, \mathcal{F})$. Then $\{X_n\}_{n\geq 1}$* ***converges in distribution*** *to a random variable $X : \Omega \to E$ if*

$$\mathbb{P}_{X_n} \to \mathbb{P}_X$$

where $\mathbb{P}_X$ is the probability measure of X. We then write $X_n \xrightarrow{d} X$.

Definition 2.3.6. *Let $X_1 : \Omega \to E$ be a random variable with a corresponding probability measure $\mathbb{P}_{X_1}$ on the measurable space $(\Omega, \mathcal{F})$. Then X_1 is* ***equal in distribution*** *to a random variable $X_2 : \Omega \to E$ if*

$$\mathbb{P}_{X_1} = \mathbb{P}_{X_2}$$

where $\mathbb{P}_{X_2}$ is the probability measure of X_2. We then write $X_1 \stackrel{d}{=} X_2$ or $X_1 \sim X_2$.

2.4 The Skorokhod Space

The stochastic processes we will be discussing in this thesis will not necessarily be continuous as they may contain jumps. Thus it is necessary to consider the Skorokhod space, which is the space of such processes, and its associated theory in order to examine the convergence of stochastic processes with jumps.

Definition 2.4.1. *The* ***Skorokhod Space*** $\mathbb{D}(\mathcal{I}, \mathbb{R}^d)$ *is the space of all* $\mathbb{R}^d$*-valued càdlàg functions defined on a subinterval* $\mathcal{I}$ *of the real line.*

In order to properly discuss convergence in the Skorokhod space we will assign it a metric. Let us first consider the simplest case of the Skorokhod space, $\mathbb{D}([0,1], \mathbb{R})$, and two functions $f_1(t)$ and $f_2(t)$ which have jumps of size one at $t = 1/2$ and $t = 1/2 + \epsilon$ respectively, but are identical otherwise. If we consider the limit as $\epsilon \to 0$, the traditional uniform metric $||x_1 - x_2||$ defined in terms of the uniform norm $||x|| := \sup_{0 \leq t \leq 1}\{|x(t)|\}$ will always equal 1. Despite the two functions becoming more similar as ϵ gets closer to 0 the metric gives no indication of the increased similarity between the two functions. The uniform metric essentially allows for functions with small differences in space to be considered close but not small differences in timing. This shortcoming of the uniform metric in the Skorokhod space motivates the use of the J_1 metric.

Definition 2.4.2. *Let* Λ *be the set of strictly increasing functions* λ *mapping the domain* $[0,1]$ *onto itself, such that both* λ *and its inverse* λ^{-1} *are continuous. Furthermore, let* $\lambda(0) = 0$ *and* $\lambda(1) = 1$*. Then, for* $x_1, x_2 \in \mathbb{D}$*, the standard* $\boldsymbol{J_1}$ ***metric*** *on* $\mathbb{D}([0,1], \mathbb{R})$ *is*

$$d_{J_1}(x_1, x_2) := \inf_{\lambda \in \Lambda}\{||x_1 \circ \lambda - x_2|| \vee ||\lambda - e||\}$$

where e *is the identity map on* $[0,1]$*.*

Intuitively Λ is the set of all the possible time distortions, thus $x_1(\lambda(t))$ can be thought of the function x_1 after a distortion of time. Using this intuition $||x_1 \circ \lambda - x_2||$ measures the difference between two functions after one has been distorted by λ, and $||\lambda - e||$ measures the size of the time distortion. Convergence is then achieved if both norms tend to zero.

In the context of our earlier example regarding $f_1(t)$ and $f_2(t)$ it is clear that $f_2(t)$ requires only an infinitesimally small distortion of time as $\epsilon \to 0$ in order for it to become identical to $f_1(t)$. Thus $f_2(t) \to f_1(t)$ as $\epsilon \to 0$ if we are equipped with the J_1 metric. It should now be easy to see that the J_1 metric allows for functions with both small differences in space and small differences in timing to be considered close.

For the rest of this thesis, we will always assume that the domain is the interval $[0, \infty)$ and the range is $\mathbb{R}$ as the stochastic processes we are concerned with are in this form. Thus we will simply write $\mathbb{D}$ to be the Skorokhod space of $\mathbb{R}$ valued functions defined on $\mathbb{R}^+$. do i need to talk about the equivalent metric which makes $\mathbb{D}$ complete

2.5 Central Limit Theorems

In the classical setting the Central Limit Theorem refers to the convergence of a sum of independent and identically distributed (i.i.d.) random variables with finite second moments towards a normal distribution. However in this thesis we will be operating in a more general setting and it is thus imperative that we define the terminology we will use throughout this thesis to avoid confusion.

Definition 2.5.1. *Consider a sequence $\{X_n : n \geq 1\}$ of $\mathbb{R}$ valued random variables and form the associated **partial sums***

$$S_n := X_1 + \cdots + X_n, \quad n \geq 1,$$

*with $S_0 = 0$. Then the associated **normalised partial sums** are*

$$\frac{S_n - mn}{b_n}$$

*where $\{b_n : n \geq 1\}$ is a sequence of constants and m is a constant. We then say that $\{X_n\}$ or $\{S_n\}$ obeys a **central limit theorem** if there exist a sequence of constants $\{b_n : n \geq 1\}$, a constant m and a nondegenerate random variable S such that there is the following convergence in distribution*

$$\frac{S_n - mn}{b_n} \xrightarrow{d} S \quad in \quad \mathbb{R}. \tag{2.5.1}$$

*We then call m the **translation scaling constant** and $\{b_n\}$ the **space scaling sequence**.*

Definition 2.5.2. *If* $\{S_n = X_1 + \cdots + X_n : n \geq 1\}$ *is a sequence of partial sums, then the associated* ***partial sum processes*** *are formed by letting*

$$S_{\lfloor nt \rfloor} := X_1 + \cdots + X_{\lfloor nt \rfloor}.$$

Definition 2.5.3. *A sequence of* ***normalised partial sum processes*** *in* $\mathbb{D}$ *associated with 2.5.1 is formed by letting*

$$\boldsymbol{S}_n(t) := \frac{S_{\lfloor nt \rfloor} - mnt}{b_n}, \quad t \geq 0, \tag{2.5.2}$$

where $\lfloor t \rfloor$ *denotes the largest integer not larger than* t*. We then say* $\{X_n\}$, $\{S_n\}$ *or* $\{\boldsymbol{S}_n\}$ *obeys a* ***functional central limit theorem*** *if there exists a proper stochastic process* $\boldsymbol{S} := \{\boldsymbol{S}(t) : t \geq 0\}$ *with sample paths in* $\mathbb{D}$ *such that*

$$\boldsymbol{S}_n \xrightarrow{d} \boldsymbol{S} \quad in \quad \mathbb{D}$$

with respect to a metric on $\mathbb{D}$*. We then call* $\boldsymbol{S}$ *the* ***stochastic process limit*** *of* $\boldsymbol{S_n}$.

Definition 2.5.4. *Let* $\{X_n : n \geq 1\}$ *be a sequence of* $\mathbb{R}$ *valued i.i.d. random variables with mean* μ *and finite variance* σ^2*. Then the* ***Classical Central Limit Theorem*** *states that*

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z \sim N(0,1) \text{ as } n \to \infty,$$

where $\bar{X}_n = (\sum_{i=1}^n X_i)/n$.

In this case we can see that the classical central limit theorem is a specific central limit theorem with the translation scaling constant $m = \mu$ and the space scaling sequence is $\{b_n = \sigma\sqrt{n} : n \geq 1\}$. It is hence obvious that the translation scaling constant and the space scaling sequence adjust for the drift and deviation of the partial sums to ensure convergence. We will look at the functional extension of the classical central limit theorem closely in Section 3.2.

Chapter 3

Limit Theorems

The aim of this chapter is to present a stochastic process limit of a bursty CTRM process. We will first rigorously define a bursty CTRM and then provide an analysis of some well known scaling limit theorems, eventually arriving at the stochastic process limit for a bursty CTRM process. We then use results for the stochastic process limit found in the literature (see [MS08]) to determine the distribution of a bursty CTRM.

3.1 Problem Formulation

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 thesis 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, and many of the results we present for the CTRM have been derived as extensions to similar results for the CTRW.

Definition 3.1.1. *Assume identically and independently distributed (i.i.d.) pairs of random variables* (J_k, W_k), $k = 1, 2, \ldots$ *where* $W_k > 0$ *represents the inter-arrival times of certain events and* $J_k \in \mathbb{R}$ *the corresponding event magnitudes. Now write*

$$S(n) = \sum_{i=1}^{n} W_i \tag{3.1.1}$$

as the sum of the first n *inter-arrival times. Also write*

$$N(t) = \max\{n \in \mathbb{N} : S(n) \leq t\} \tag{3.1.2}$$

for the renewal process associated with the W_k. Then the process

$$V(t) = M(N(t)) = \bigvee_{k=1}^{N(t)} J_k = \max\{J_k : k = 1, \ldots, N(t)\}, \quad t \geq 0. \tag{3.1.3}$$

is called a CTRM (Continuous Time Random Maxima process), where the maximum of the empty set is set to 0.

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 the largest magnitude observed until 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

$$\tilde{M}(t) = \bigvee_{k=1}^{N(t)+1} J_k, \quad t \geq 0. \tag{3.1.4}$$

Finally, the model is called *coupled* when W_k and J_k are not independent. In this thesis 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 $M(t)$.

expand. It is for this reason that we choose to make inference on the distribution of the following quantities.

Definition 3.1.2. *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

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

These quantities are chosen as it is natural to be concerned with both the size and the timing of events of large magnitude. Possessing a strong understanding of these large events will aid in managing the consequences associated with extreme natural phenomena, financial loss and human behaviour. reword this

Lemma 3.1.3. *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 rigorize this

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

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

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

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

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

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. □

The distribution of the exceedance is hence, as expected, simply the conditional distribution of a magnitude J_k given $J_k > \ell$. We will look at the distributions of exceedances and exceedance times in further detail in the next few sections.

3.2 Brownian Motion as a Stochastic Process Limit

A stochastic process limit is the limit of a sequence of stochastic processes. Before we examine the stochastic process limits that we will utilise in this thesis it will be illustrative to analyse one of the most well known examples of a stochastic process limit - Brownian Motion. One of the methods used to derive Brownian motion is devised by trying to seek a continuous time representation of the discrete time random walk. Recall that a discrete time random walk is defined in the following manner.

Definition 3.2.1. *Let $\{U_i\}_{1\leq i\leq n}$ be i.i.d. U(0,1) random variables. Write the partial sums at time k as*

$$S_k = \sum_{i=1}^{k} U_i,$$

then successive partial sums form a ***random walk*** *with S_n being the position after n steps.*

The general method of achieving a continuous limit of a discrete function is to consider a continuous analogue of the sum S_n as $n \to \infty$ and then rescaling to ensure the continuous analogue remains finite. We can then appeal to the Classical Central Limit Theorem which gives us the discrete limit

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} Z \sim N(0,1) \text{ as } n \to \infty \tag{3.2.1}$$

where $\mu := \mathbb{E}[U_k] = 1/2$ and $\sigma^2 := \mathrm{Var}(U_k) = 1/12$. The random walk only accepts integer arguments so in order to extend this discrete limit to a continuous limit we first consider two different continuous time representations of a discrete time random walk. One obvious way of forming a process from the random walk is to perform linear interpolation between each discrete point. Thus, we can consider

$$\tilde{S}(t) := (t - \lfloor t \rfloor)U_{\lfloor t \rfloor + 1} + S_{\lfloor t \rfloor} \quad \text{for all } t \geq 0, \tag{3.2.2}$$

or more simply, we can consider the step function

$$S_{\lfloor t \rfloor} = U_1 + \cdots + U_{\lfloor t \rfloor} \quad \text{for all } t \geq 0. \tag{3.2.3}$$

As it turns out for large values of t both processes are equivalent, see [Whi01] for more details. Although the interpolation function is more intuitive, the step function is easier to work with and so we only utilise the step function from now on. We now add a time scaling constant c and consider the partial sum processes

$$S_{\lfloor ct \rfloor} = U_1 + \cdots + U_{\lfloor ct \rfloor} \quad \text{for all } t \geq 0. \tag{3.2.4}$$

Whereas in 3.2.3 a step can only occur at the end of every time period, the inclusion of c in 3.2.4 allows us to adjust how often steps occur. For example, if we multiply c by 60, then steps that were occurring every minute now occur every second. Clearly as c approaches infinity steps start to occur continuously. Now that we have a continuous analogue of the discrete random walk we wish to find the limiting process. Following 3.2.1 we would expect that the following functional central limit theorem holds

$$\frac{S_{\lfloor ct \rfloor} - \lfloor ct \rfloor \mu}{\sqrt{c\sigma^2}} \xrightarrow{d} B_t \sim N(0,t) \text{ as } c \to \infty, \tag{3.2.5}$$

where B_t is the standard Brownian Motion. An easy way to check that the above weak convergence is correct is to consider the Fourier transforms of both sides.

Definition 3.2.2. *Let $Y : \Omega \to E$ be a random variable with probability density function $f_Y(x)$. Then the* ***Fourier Transform*** *of $f_Y(x)$ is*

$$\hat{f}_Y(s) = \mathbb{E}[e^{-isY}] = \int_\Omega e^{-isx} f_Y(x) dx.$$

When the first two moments are finite we can use the Taylor series expansion of $e^z = 1 + z + z^2/2! + ...$ to see that

$$\begin{aligned}
\hat{f}_Y(s) &= \int_\Omega \left(1 - isx + \frac{1}{2!}(-isx)^2 + \cdots \right) f_Y(x) dx && (3.2.6)\\
&= \int_\Omega f_Y(x) dx - is \int_\Omega x f_Y(x) dx - \frac{s^2}{2} \int_\Omega x^2 f_Y(x) dx + \cdots && (3.2.7)\\
&= 1 - is\mathbb{E}[Y] - \frac{s^2}{2}\mathbb{E}[Y^2] + o(s^2), && (3.2.8)
\end{aligned}$$

where $o(s^2)$ is a function that approaches zero faster than s^2 as $s \to 0$.

Suppose $Y \sim Y_k = U_k - \mu$ then

$$\hat{f}_Y(s) = 1 - \frac{s^2}{2}\sigma^2 + o(s^2), \tag{3.2.9}$$

since $\mathbb{E}[Y] = 0$, and $\mathbb{E}[Y^2] = \mathrm{Var}(Y) + \mathbb{E}[Y]^2 = \sigma^2$. We now have the necessary tools to calculate the Fourier transform of the rescaled $S_{\lfloor ct \rfloor}$ defined in 3.2.5. Using the fact that the Fourier transform of a sum is simply the product of the individual Fourier transforms we then have

$$\begin{aligned}
\mathbb{E}\left[\exp\left(-is\frac{S_{\lfloor ct \rfloor} - \lfloor ct \rfloor \mu}{\sqrt{c\sigma^2}}\right)\right] &= \mathbb{E}\left[\exp\left(-is\frac{Y_1 + \cdots + Y_{\lfloor ct \rfloor}}{\sqrt{c\sigma^2}}\right)\right] && (3.2.10)\\
&= \mathbb{E}\left[\exp\left(\frac{-isY}{\sqrt{c\sigma^2}}\right)\right]^{\lfloor ct \rfloor} && (3.2.11)\\
&= \hat{f}_Y\left(\frac{s}{\sqrt{c\sigma^2}}\right)^{\lfloor ct \rfloor}. && (3.2.12)
\end{aligned}$$

Given that $(1 + (r/n) + o(n^{-1}))^n \to e^r$ as $n \to \infty$ for any $r \in \mathbb{R}$ we then have

$$\begin{aligned}
\mathbb{E}\left[\exp\left(-is\frac{S_{\lfloor ct \rfloor} - \lfloor ct \rfloor \mu}{\sqrt{c\sigma^2}}\right)\right] &= \hat{f}_Y\left(\frac{s}{\sqrt{c\sigma^2}}\right)^{\lfloor ct \rfloor} \\
&= \left(1 - \frac{s^2}{2c} + o(c^{-1})\right)^{\lfloor ct \rfloor} && (3.2.13)\\
&= \left[\left(1 - \frac{s^2}{2c} + o(c^{-1})\right)^c\right]^{\frac{\lfloor ct \rfloor}{c}} && (3.2.14)\\
&\to e^{-ts^2/2} \quad \text{as } c \to \infty, && (3.2.15)
\end{aligned}$$

since $\lfloor ct \rfloor / c \to t$ as $c \to \infty$.

We now refer to the Lévy Continuity Theorem to prove the weak convergence outlined in 3.2.5.

Definition 3.2.3. *[MS01]* ***Lévy Continuity Theorem****. If X_n, X are random variables on $\mathbb{R}$, then $X_n \xrightarrow{d} X$ implies that $\hat{f}_{X_n}(s) \to \hat{f}_X(s)$ for each $s \in \mathbb{R}$, uniformly on compact subsets. Conversely, if X_n is a sequence of random variables such that $\hat{f}_{X_n}(s) \to \hat{f}_X(s)$ for each $s \in \mathbb{R}$, and the limit is continuous at $s = 0$, then $\hat{f}_X(s)$ is the Fourier transform of some random variable X, and $X_n \xrightarrow{d} X$.*

This theorem essentially states that $X_n \xrightarrow{d} X$ if and only if $\hat{f}_{X_n}(s) \to \hat{f}_X(s)$. We previously showed that the limit of the Fourier transform of the rescaled $S_{\lfloor ct \rfloor}$ is $e^{-ts^2/2}$ which is also the Fourier transform of a $N(0,t)$ random variable. Thus we must have the result outlined in 3.2.5.

$$\frac{S_{\lfloor ct \rfloor} - \lfloor ct \rfloor \mu}{\sqrt{c\sigma^2}} = \frac{\sqrt{\lfloor ct \rfloor}}{\sqrt{\lfloor ct \rfloor}} \frac{S_{\lfloor ct \rfloor} - \lfloor ct \rfloor \mu}{\sqrt{c\sigma^2}} \to \sqrt{t} B_1 \stackrel{d}{=} B_t$$

by (3.2.1).

Now that we have shown how one can derive Brownian Motion as a stochastic process limit of the rescaled $S_{\lfloor ct \rfloor}$, the stochastic process limits presented in later sections will be more accessible.

3.3 Stable Distributions

In order for our inter-arrival waiting times W_k to be modelled correctly they should be heavy tailed and completely right skewed so as not to violate our initial bursty assumption. Stable distributions are an extensive class of probability distributions that allow for both skew and heavy tails and are thus a perfect distribution to describe our waiting times with. As we will find out later in this section they also have many useful properties. The class was first defined by Paul Lévy in the early 20th century but was under utilised due to the lack of closed formulas for densities and distribution functions for all but a few stable distributions (Gaussian, Cauchy and Lévy). However its use has increased since advances in computing methods expand now allow for the computation of stable densities, distribution functions and quantiles.

Definition 3.3.1. *[Nol15] A random variable X is **sum-stable** if for X_1 and X_2 independent copies of X and any positive constants a and b we have*

$$aX_1 + bX_2 \stackrel{d}{=} cX + d$$

*for some positive c and some $d \in \mathbb{R}$. We then say that X follows a **stable distribution**. Note that the symbol $\stackrel{d}{=}$ was defined in Definition 2.3.6.*

Remark 3.3.2. *Note that other texts refer to what we have just defined as **sum-stable** as just **stable**. We make this modification in order to easily differentiate it* *from the term **max-stable** which we will define in section 3.5.*

As a closed form for the density of a stable random variable cannot be found for all cases we instead choose to specify a stable random variable by its characteristic function.

Definition 3.3.3. *Let $X \sim S(\alpha, \beta, \sigma, \mu)$ be a stable random variable. Then the characteristic function $\phi_X(s) = \mathbb{E}[e^{isX}]$ is given by*

$$\phi_X(s) = \begin{cases} \exp\left(-\sigma^\alpha \, |s|^\alpha \, [1 - i\beta \operatorname{sign}(s) \tan(\frac{\pi\alpha}{2})] + i\mu s\right) & \alpha \neq 1 \\ \exp\left(-\sigma \, |s| \, [1 + i\beta \operatorname{sign}(s) \frac{2}{\pi} \log(s)] + i\mu s\right) & \alpha = 1. \end{cases} \tag{3.3.1}$$

There are over half a dozen parameterisations of a stable random variable, but we use [ST94, Def 1.1.6]. Definition 3.3.3 then implies that a stable distribution possesses four different parameters:

- An index of stability $\alpha \in (0, 2]$ which determines heavy tailedness
- A skewness parameter $\beta \in [-1, 1]$
- A scale parameter $\sigma > 0$
- A location parameter $\mu \in \mathbb{R}$

As we are using the stable distribution to model bursty waiting times we need a completely right skewed distribution with a heavy tail (infinite mean). Thus we will only consider waiting times drawn from a stable distribution with:

- Index of stability $\alpha \in (0, 1)$ to ensure heavy tailedness
- Skewness parameter $\beta = 1$ to ensure a right skewed distribution
- Unspecified scale and location parameters σ and μ

To check that the mean is infinite when $\beta \in (0, 1)$ we first see from [JW94] that if a Central Limit Theorem type argument is used we have

$$\lim_{x \to \infty} x^\beta \mathbb{P}(X > x) = C_\beta \frac{1+\alpha}{2} \gamma^\beta, \tag{3.3.2}$$

$$\lim_{x \to \infty} x^\beta \mathbb{P}(X < -x) = C_\beta \frac{1-\alpha}{2} \gamma^\beta, \tag{3.3.3}$$

where

$$C_\beta = \left(\int_0^\infty x^{-\beta} \sin(x) dx \right)^{-1} = \frac{2}{\pi} \Gamma(\beta) \sin \frac{\pi\beta}{2}.$$

When $\alpha = 1$ we have $\lim_{x \to \infty} \mathbb{P}(X < -x) = 0$, as one would expect. Since a Pareto random variable T with parameter β has a tail function $\mathbb{P}(T > t) = ct^{-\beta}$ for some

$c \in \mathbb{R}$ and $t > c^{1/\beta}$, we can then observe that the tails of a stable distribution are asymptotically equivalent to a Pareto distribution and thus exhibit power-law behaviour. Thus if we show that a Pareto distribution with $\beta \in (0, 1)$ has an infinite mean then the corresponding stable distribution also has an infinite mean. Now if $p(t)$ is the density of T we have that

$$\mathbb{E}[T] = \int_0^\infty tp(t)dt = \int_0^\infty t\frac{c\beta}{t^{\beta+1}}dt = c\beta \int_0^\infty t^{-\beta}dt = c\beta \left[\frac{t^{1-\beta}}{1-\beta}\right]_0^\infty .$$

The above expectation is clearly infinite if $0 < \beta \leq 1$, but if $\beta = 1$ our initial stable distribution reduces to a Cauchy distribution and thus we exclude that case.why ignore Cauchy?? also maybe talk about variance here

One of the reasons the stable distribution is so useful is that we can now generalise the classical Central Limit Theorem. Recall that the Central Limit Theorem says that the normalised sum of i.i.d. random variables with a finite variance converges to a normal distribution. Rewriting the above to match the notation we will use from now on we have

$$\frac{\sum_{i=1}^n X_i - a_n}{b_n} \xrightarrow{d} Z \sim N(0, 1) \text{ as } n \to \infty$$

where $a_n = \sqrt{n}\sigma/\mu$ and $b_n = \sigma\sqrt{n}$.

The Generalised Central Limit Theorem generalises the Central Limit Theorem by dropping the assumption that the variance, or even the mean, is finite. This theorem is important to us as we are assuming our waiting times W_k have an infinite mean.

Theorem 3.3.4. *[Nol15]* ***Generalised Central Limit Theorem*** *A nondegenerate random variable Z follows a stable distribution if and only if there is an i.i.d. sequence of random variables $\{X_i\}_{1 \leq i \leq n}$ and constants $a_n > 0, b_n \in \mathbb{R}$ with*

$$\frac{\sum_{i=1}^n X_i - a_n}{b_n} \xrightarrow{d} Z.$$

The Generalised Central Limit Theorem implies that the only possible non-trivial limit of the normalised sum of our waiting times is a stable distribution. If a limit exists we can then extend this limit to stochastic process convergence as in 3.2. The theorem also naturally leads to the following definition.

Definition 3.3.5. *[Nol15] A random variable X is in the* ***domain of attraction*** *of Z if there exists constants $a_n > 0, b_n \in \mathbb{R}$ with*

$$\frac{\sum_{i=1}^{n} X_i - a_n}{b_n} \xrightarrow{d} Z,$$

where $X, X_1, X_2, X_3, \ldots$ are i.i.d. We then say $X \in DOA(Z)$.

The above definition will be useful in section 3.4 where we present a stochastic process limit for the the sum of our inter-arrival waiting times.

3.4 Scaling Limit of the Sum of Waiting Times

In order to introduce more useful necessary and sufficient conditions for the Generalised Central Theorem to hold we must first define the concept of regular variation.

Definition 3.4.1. *Suppose that $R : [A, \infty) \to (0, \infty)$ is Borel measurable, for some $A > 0$. We say that $R(x)$* ***varies regularly*** *or is* ***regularly varying*** *with index ρ, and we write $R \in RV(\rho)$, if*

$$\lim_{x \to \infty} \frac{R(\lambda x)}{R(x)} = \lambda^{\rho},$$

for all $\lambda > 0$. If $\rho = 0$ we then say $R(x)$ is ***slowly varying*** *or* ***varies slowly****.*

Remark 3.4.2. *It is easy to see that every regularly varying function R of index ρ has representation*

$$R(x) = x^{\rho} L(x)$$

where L is some slowly varying function.

Now we can state the extended limit theorem which provides us with a useful criterion that ensures the sum of our waiting times converges to a nondegenerate random variable.

Theorem 3.4.3. *[MS12, Th. 4.5]* ***Extended Central Limit Theorem*** *If Z is sum-stable with index $0 < \alpha < 2$, then $X \in DOA(Z)$ if and only if $\mathbb{P}[|X| > x]$ is regularly varying with index $-\alpha$ and*

$$\lim_{x\to\infty} \frac{\mathbb{P}[X > x]}{\mathbb{P}[|X| > x]} = p \quad \textit{for some } 0 \le p \le 1. \tag{3.4.1}$$

Thus the Extended Central Limit Theorem in conjunction with Definition 3.3.5 tells us that the normalized sum of random variables with an infinite mean will converge in distribution to a nondegenerate random variable as long as the condition outlined in 3.4.1 holds and the tail is regularly varying.

As our waiting times W_k are strictly greater than zero, they are completely right skewed and condition 3.4.1 always holds. So if we assume a heavy-tailed distribution for the waiting times such that the tail function $\overline{F}_W := 1 - F_W$ of the CDF of W is regularly varying, i.e.

$$t \mapsto \overline{F}_W(x) \in RV(-\alpha), \quad \alpha \in (0,1) \tag{3.4.2}$$

meaning that [Sen76, MS01],3.4.1

$$\lim_{x\to\infty} \frac{\overline{F}_W(\lambda x)}{\overline{F}_W(x)} = \lambda^{-\alpha}, \quad \lambda > 0.$$

Then by Theorem 3.4.3 this is equivalent to W_k being in the (sum-) domain of attraction of a positively skewed stable law D with stability parameter α, which means the weak convergence

$$\frac{W_1 + \ldots + W_n}{b(n)} \xrightarrow{d} D, \quad n \to \infty. \tag{3.4.3}$$

Note that the translation scaling constant $a(n)$ is not required when $\alpha \in (0,1)$ as the mean of W_k is then infinite, see [Whi01] for more details. We also have that $b(n) \in RV(1/\alpha)$ [MS12, Prop 4.15]. In fact if we specify W_k to be Pareto distributed such that $\overline{F}_W(x) = Cx^{-\alpha}$ for some $C > 0$, for all $x > C^{1/\alpha}$ and $0 < \alpha < 1$. Then $\overline{F}_W(x)$ is clearly regularly varying with index $-\alpha$ and by [MS12, Th 3.37] we have the more specific result

$$\frac{W_1 + \ldots + W_n}{n^{1/\alpha}} \xrightarrow{d} Y, \quad n \to \infty. \tag{3.4.4}$$

where Y is stably distributed and has the characteristic function

$$\phi_Y(s) = \mathbb{E}[e^{isY}] = \exp[-C\Gamma(1-\alpha)(-is)^\alpha], \tag{3.4.5}$$

where $\Gamma(x)$ is the gamma function. Now that the convergence for the sum of our waiting times has been established, in a manner similar to how we arrived at the stochastic process limit of a discrete random walk (3.2.5), we can extend this result to a stochastic process limit.

Theorem 3.4.4. *[MS12, Th 3.41] Suppose W_k are i.i.d. such that $\overline{F}_W(x) = Cx^{-\alpha}$ for some $C > 0$, for all $x > C^{1/\alpha}$ and $0 < \alpha < 1$. Then*

$$c^{-1/\alpha} \sum_{j=1}^{\lfloor ct \rfloor} W_j \xrightarrow{d} Z(t), \quad c \to \infty, \tag{3.4.6}$$

for all $t > 0$, where

$$\phi_{Z_t}(s) = \mathbb{E}[e^{isZ_t}] = \exp[-Ct\Gamma(1-\alpha)(-is)^\alpha], \tag{3.4.7}$$

writing $Z(t) := Z_t$.

Proof. From [MS12, Th 3.39] and 3.4.4 we have $n^{-1/\alpha}(W_1 + \ldots + W_n) \xrightarrow{d} Y$ as $n \to \infty$, and the limit Y has characteristic function $\phi_Y(s)$ defined in 3.4.5. Let $\phi_{W_n}(s) = \mathbb{E}[\exp(n^{-1/\alpha}W_1)]$ be the characteristic function of a scaled waiting time. By Lévy's Continuity Theorem (3.2.3) we have that $\phi_{W_n}(s)^n \to \phi_Y(s)$. The properties of a characteristic function then tell us that the characteristic function of the scaled partial sum $n^{-1/\alpha} \sum_{j=1}^{\lfloor nt \rfloor} W_j$ is

$$\phi_{W_n}(s)^{\lfloor nt \rfloor} = (\phi_{W_n}(s)^n)^{\frac{\lfloor nt \rfloor}{n}} \to \phi_Y(s)^t, \quad n \to \infty,$$

for any $t > 0$ because $\frac{\lfloor nt \rfloor}{n} \to t$ as $n \to \infty$. Once again by Lévy's Continuity Theorem we have that the limit Z_t has characteristic function

$$\phi_{Z_t}(s) = \phi_Y(s)^t = \exp[-Ct\Gamma(1-\alpha)(-is)^\alpha].$$

Notation issue switching from n to c when considering stochastic process limit ☐

Furthermore if we drop the assumption that W_k is Pareto distributed but retain the assumption that they strictly greater than zero then we get a more general, functional limit theorem. However before we do that we must introduce *stable subordinators* which are a class of stochastic processes.

Definition 3.4.5. *A **stable subordinator** $D(t)$ is a nondecreasing Lévy process such that the increments have right skewed stable laws, in particular*

$$D(t+s) - D(t) \stackrel{d}{=} t^{1/\alpha} S(\alpha, 1, \sigma, 0)) \stackrel{d}{=} S(\alpha, 1, t^{1/\alpha}\sigma, 0).$$

α-stable subordinators are stable subordinators with stability parameter α.

Theorem 3.4.6. *[Whi01, Th. 4.5.3] If W_k obey a central limit theorem such that*

$$\frac{W_1 + \ldots + W_n - a(n)}{b(n)} \xrightarrow{d} D, \quad n \to \infty, \tag{3.4.8}$$

where $D \sim S(\alpha, 1, \sigma, 0)$. Then there is convergence in distribution for the associated normalised partial sum process

$$\frac{S_{\lfloor ct \rfloor} - a(c)t}{b(c)} \xrightarrow{d} D(t), \quad c \to \infty, \tag{3.4.9}$$

in the Skorokhod space $\mathbb{D}$ with respect to the J_1 metric, where $D(t)$ is a α-stable subordinator and $S_{\lfloor ct \rfloor} = W_1 + \ldots + W_{\lfloor ct \rfloor}$.

Note that the translation scaling constant $a(c)$ is not required when $\alpha \in (0, 1)$. We can also adjust $b(c)$ such that the stable subordinator limit is a *unit* stable subordinator with respect to the Laplace transform, defined via

$$\mathbf{E}[e^{-\lambda D(t)}] = e^{-t\lambda^\alpha}. \tag{3.4.10}$$

This theorem in conjunction with the extended central limit theorem 3.4.3 thus ensures that we have an α-stable subordinator as the stochastic process limit of the waiting times as long as W_k are strictly positive and regularly varying with index $-\alpha$.

Now that we have a stochastic limit process for the sum of the waiting times we can investigate the limiting behaviour of the renewal process in equation 3.1.2. Recall that the renewal process is the process which counts the amount of waiting times that have occurred up until time t. If W_k obeys the central limit theorem outlined in Equation 3.4.8 then the scaling limit of the renewal process is then

[MS04]

$$\tilde{b}(c)^{-1}N(ct) \xrightarrow{d} E(t), \quad c \to \infty \tag{3.4.11}$$

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

$$E(t) = \inf\{r : D(r) > t\}, \quad t \geq 0, \tag{3.4.12}$$

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

$$b(\tilde{b}(c)) \sim c \sim \tilde{b}(b(c)) \tag{3.4.13}$$

where a $\sim$ symbol indicates that the quotient of both sides converges to 1 as $c \to \infty$. Note that $\tilde{b}(c) \in RV(\alpha)$ as $b(c) \in RV(1/\alpha)$ [MS12, Prop 4.15].

A subordinator is a stochastic process that determines the evolution of time within another stochastic process, which is referred to as the subordinated stochastic process. The inverse stable subordinator, $E(t)$, is a subordinator itself (as the inverse of a nondecreasing function is nondecreasing) and it governs the temporal dynamics of the scaling limit of the CTRM process $M(t)$, see Theorem 3.6.1. 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 then 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 ($\alpha \in (0,1)$ and a scale parameter) the inverse stable subordinator hence models scaling limits of heavy-tailed waiting times parsimoniously.

3.5 Scaling Limit of the Maximum of Event Magnitudes

The aim of this section is to develop a stochastic process limit theorem for the partial maxima of our event magnitudes J_k. The partial maxima is defined as

$$M_n := \bigvee_{i=1}^{n} J_i. \tag{3.5.1}$$

Classical extreme value theory proves to be quite useful in solving this problem as a large portion of the extreme value literature focuses on the behaviour of the partial maxima M_n. Classical extreme value theory was first developed as a solution to a problem that arises when dealing with partial maxima. Assume our event magnitudes J_k have common distribution function $F(x)$, then the distribution of the partial maxima can easily be calculated as

$$\begin{aligned}\mathbb{P}(M_n \leq x) &= \mathbb{P}\left(\bigvee_{i=1}^{n} J_i \leq x\right) \\ &= \mathbb{P}(J_1 \leq x) \times \mathbb{P}(J_2 \leq x) \times \ldots \times \mathbb{P}(J_n \leq x) \\ &= [F(x)]^n.\end{aligned}$$

However when when you consider the limit as $n \to \infty$ the distribution of M_n converges to a degenerate distribution. That is

$$\mathbb{P}(M_n \leq x) = \begin{cases} 1 & \text{if } x \geq x^+ \\ 0 & \text{if } x < x^+ \end{cases}$$

where x^+ is the smallest value such that $F(x) = 1$. This result is clearly of little use in describing the behaviour of M_n. This problem was overcome by considering the normalised partial sum

$$M_n^* := \frac{M_n - d_n}{a_n},$$

where a_n is a sequence of positive constants and $d_n \in \mathbb{R}$. These constants could then be picked such that the location and scale of M_n^* stabilise as $n \to \infty$. This idea was then encapsulated in the following theorem.

Theorem 3.5.1. *[Col01]* ***Fisher-Tippet-Gnendenko Theorem*** *Let $J_1, J_2, \ldots, J_n$ be i.i.d. random variables with distribution function $F(x)$ and partial maxima M_n. If there exists a sequence of positive constants a_n and $d_n \in \mathbb{R}$ such that*

$$\mathbb{P}\left\{\frac{M_n - d_n}{a_n} \leq x\right\} = F^n(a_n x + d_n) \xrightarrow{d} G(x) \quad \text{as } n \to \infty,$$

where $G(x)$ is a non-degenerate distribution function, then G must follow one of three distributions:

Type I (Gumbel)

$$G(x) = \exp\left\{-\exp\left[-\left(\frac{x-b}{a}\right)\right]\right\}, \forall x$$

Type II (Fréchet)

$$G(x) = \begin{cases} 0 & x \leq b \\ \exp\left\{-\left(\frac{x-b}{a}\right)^{-\alpha}\right\} & x > b \end{cases}$$

Type III (Weibull)

$$G(x) = \begin{cases} \exp\left\{-\left[-\left(\frac{x-b}{a}\right)^{\alpha}\right]\right\} & x < b \\ 1 & x \geq b \end{cases}$$

for scale parameters $a > 0$, location parameter b, and in the case of families II and III, shape parameter $\alpha > 0$.

The Fisher-Tippet-Gnendenko Theorem was first introduced by Fisher and Tippet [?] and was extended by Gnedenko [?]. It implies that the only possible non-trivial limit of the normalised partial maxima of our event magnitudes is either a Gumbel, Fréchet or Weibull distribution. Despite the fact that this theorem is expressed in terms of the distribution functions rather than random variable one can see how it is an analogue of the classical central limit theorem in a max sense as opposed to a sum sense.

It turns out that all three distributions can be expressed in terms of a single parameterisation. This parameterisation is known as the generalized extreme value (GEV) family of distributions. It is thus more convenient to express the above central limit theorem in terms of the GEV family of distributions.

Theorem 3.5.2. *[Col01] Let $J_1, J_2, \ldots, J_n$ be i.i.d. random variables with distribution function $F(x)$ and partial maxima M_n. If there exists a sequence of constants $a_n > 0$ and $d_n \in \mathbb{R}$ such that*

$$\mathbb{P}\left\{\frac{M_n - d_n}{a_n} \leq x\right\} \xrightarrow{d} G(x) \quad \text{as } n \to \infty,$$

where $G(x)$ is a non-degenerate distribution function, then $G(x)$ is a member of the GEV family

$$G(x) = \exp\left\{-\left[1 + \zeta\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\zeta}\right\} \tag{3.5.2}$$

defined on $\{x : 1 + \zeta(x-\mu)/\sigma > 0\}$*, where* μ *is finite,* $\sigma > 0$ *and* ζ *is finite. We can then say* $F(x)$ *lies in the* ***max-domain of attraction*** *of* $G(x)$.

In the above parameterisation the Gumbel family is achieved when $\zeta \to 0$, the Fréchet family corresponds to the case when $\zeta > 0$ and the Weibull family corresponds to the case when $\zeta < 0$. We will now use the following definition to refer to distributions which remain as the same distribution (up to location and scale constants) after taking the maximum of multiple i.i.d. copies.

Definition 3.5.3. *[BR77] A distribution G(x) is said to be* ***max-stable*** *if for every* $n \in \mathbb{N}$ *there are constants* $a_n > 0$ *and* d_n *such that*

$$G^n(a_n x + d_n) = G(x).$$

Resnick and Davis in [DR84] were able to show that this definition is an equivalent characterisation of a GEV random variable. They proved a distribution is max-stable iff it is a GEV distribution. Note that this result can be seen as a max counterpart of Theorem 3.3.4. Now that we have stated the max analogue of the central limit theorem and its associated terminology we attempt to seek a more general functional limit theorem. But before we do that we must provide the definition of an *extremal process* which is the stochastic process extension of a GEV random variable.

Definition 3.5.4. *Let X(t) be a stochastic process. X(t) is an* ***extremal process*** *if for* $d \geq 1$, $x_i \in \mathbb{R}$ *and the times* $0 = t_0 < t_1 < \ldots < t_d$ *we have*

$$\mathbb{P}(X(t_i) \leq x_i, 1 \leq i \leq d) = F(\wedge_{i=1}^d x_i)^{t_1} F(\wedge_{i=2}^d x_i)^{t_2 - t_1} \ldots F(x_d)^{t_d - t_{d-1}},$$

where F(x) is a GEV distribution. We then say X(t) is an extremal process ***generated*** *by F(x). Equivalently if* J_i *are i.i.d. random variables then X(t) is an extremal process if*

$$X(t_1); X(t_2); \ldots X(t_d);$$

has the same joint distribution as

$$J_1; max(J_1, J_2); \ldots; max(J_1, J_2, \ldots, J_d),$$

where $0 \leq t_1 < t_2 < \ldots < t_d$*, and* $\mathbb{P}(J_i < x) = \mathbb{P}(X(J_i - J_{i-1})$ *for* $i = 1, 2, \ldots, d$.

In 1964 Lamperti [Lam64] was able to show that the same conditions that are required for a partial maxima to be nondegenerate can be used to show that the partial maxima-process will converge to an extremal process.

Theorem 3.5.5. *[Lam64, Th 3.2] (see also [Res74, Th 2]) Let $F(x)$ be the CDF of J_i. Now suppose there exists constants $a(n) > 0$ and $d(n)$ such that,*

$$\mathbb{P}\left(\frac{M_n - d_n}{a_n} \leq x\right) = F^n\left(a_n x + d_n\right) \xrightarrow{d} G(x).$$

Then $G(x)$ must be a member of the GEV family. Now define the partial maxima-process as

$$M_{\lfloor t \rfloor} := \begin{cases} \bigvee_{i=1}^{\lfloor t \rfloor} J_i, & t \geq 1 \\ J_1, & 0 < t < 1. \end{cases}$$

Then

$$\frac{M_{\lfloor ct \rfloor} - d(c)}{a(c)} \xrightarrow[c \to \infty]{J_1} A(t),$$

where $\{A(t)\}_{t \geq 0}$ is an extremal process generated by G.

Just as we sought random variables that remained nondegenerate after addition to model our waiting times, we now seek random variables that remain nondegenerate after taking the maximum to model our event magnitudes. By Theorem 3.5.5 this will ensure that the normalised partial maxima-process will converge to an extremal process. From [BGST06] we know that any probability distribution with a continuous distribution $F(x)$ on $\mathbb{R}$ lies in the max-domain of attraction of a GEV distribution $G(x)$. Furthermore the distribution of the largest magnitudes is well approximated by a GEV distribution $G(x)$ if the number of events is large [Col01]. Thus we simply require the event magnitudes to be drawn from a continuous distribution, whereas in the sum-stable case we required the waiting times to have regularly varying tails.

3.6 Scaling Limit of CTRM

In the previous two sections we outlined the necessary conditions for our waiting times W_k to be in the domain of attraction of a sum-stable random variable and our event magnitudes J_k to be in the domain of attraction of a max-stable (GEV) random variable. We also discussed how the same conditions then allowed for the partial sum-process $S_{\lfloor ct \rfloor}$ and partial max-process $M_{\lfloor ct \rfloor}$ (both normalised appropriately) to converge to a stable subordinator and extremal process respectively. More precisely if our waiting times W_k have tail function $\overline{F}_W \in RV(-\alpha)$ for $\alpha \in (0, 1)$, and if our event magnitudes J_k are i.i.d. and drawn from a continuous probability distribution $F(x)$ in the max-domain of attraction of $G(x)$, then we have a limit theorem for our partial sum-process,

$$b(c)^{-1}S_{\lfloor ct\rfloor} \xrightarrow{d} D(t), \quad c \to \infty, \tag{3.6.1}$$

a limit theorem for our renewal process,

$$N(ct)/\tilde{b}(c) \xrightarrow{d} E(t), \quad c \to \infty, \tag{3.6.2}$$

and a limit theorem for our partial maxima-process,

$$\frac{M_{\lfloor ct\rfloor} - d(c)}{a(c)} \xrightarrow{d} A(t), \quad c \to \infty. \tag{3.6.3}$$

Recall that $b(c) \in RV(1/\alpha)$ could be chosen such that $D(t)$ was a unit stable subordinator with Laplace transform $\mathbb{E}[e^{-\lambda D(t)}] = e^{-t\lambda^\alpha}$ (3.4.10), $\tilde{b}(c) \in RV(\alpha)$ was the asymptotic inverse of $b(c)$ (3.4.13), $E(t) = \inf\{r : D(r) > t\}$ is an inverse stable subordinator (3.4.12), and $A(t)$ is the extremal process generated by $G(x)$ (3.5.5).

Now that we have summarised all the limit theorems presented so far we are in a position to outline the functional limit theorem for the CTRM process that Meerschaert and Stoev were able to prove in 2008.

Theorem 3.6.1. *[MS08] Let (W_i, J_i) be a sequence of i.i.d $\mathbb{R}^+ \times \mathbb{R}$ random vectors such that the limits in equations 3.6.1,3.6.2 and 3.6.3 hold. Then,*

$$\frac{M(N(ct)) - d(\tilde{b}(c))}{a(\tilde{b}(c))} \xrightarrow{d} A(E(t)), \quad c \to \infty,$$

where $A(E(t))$ is a subordination of the extremal process $A(t)$ by the inverse stable subordinator process $E(t)$.

Recall that $M(N(t)) = V(t)$ is our CTRM process. Now that we have its scaling limit we can use results derived for the extremal process subordinated to the inverse stable subordinator to derive the distributions of the exceedances and their durations.

Chapter 4

Distribution of CTRM

In this chapter we derive the distribution of a CTRM process if the inter arrival waiting times that drive the process exhibit a bursty nature.

4.1 Distribution of Exceedances

4.2 Distribution of Exceedance Durations

An Overview of Generalized Gamma MittagLeffler Model and Its Applications

Chapter 5

Statistical Inference

5.1 Log Moment Estimator for Mittag-Leffler Random Variables

5.2 Maximum Likelihood Estimator for Generalised Pareto Random Variables

5.3 Stability of Parameter Estimates

5.4 Goodness of Fit

We will test Goodness of Fit as Mainardi et al with the reduced chi squared test

Chapter 6

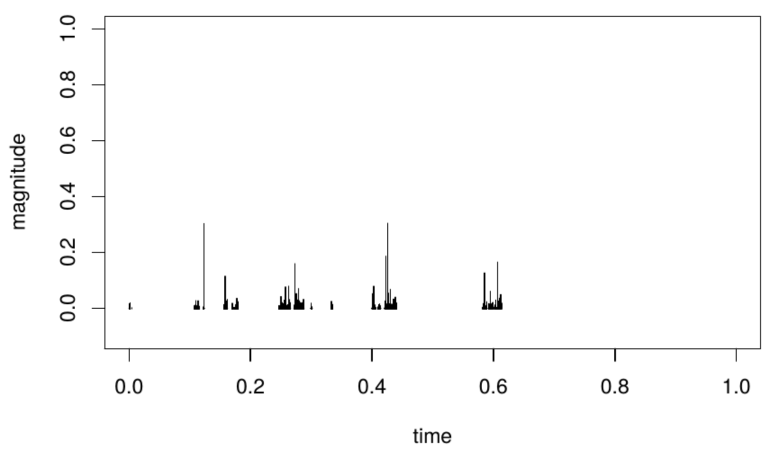
Model Fitting

The accuracy of the Mittag-Leffler log moment estimator and the Generalised Pareto MLE has been studied in [Cah13] and [?] respectively. However it is yet to be confirmed that these estimation methods work for exceedances and exceedance durations. Thus before we apply our inference methods to actual data, we will check that we can correctly estimate the theoretical Mittag-Leffler and Generalised Pareto parameters derived from a simulated bursty process. If we are able to accurately estimate these parameters we can then proceed to apply the estimation methods to actual data, confident in knowing that the actual parameters will be estimated.

6.1 Model Fit of Simulated Data

We assume n i.i.d. waiting times are drawn from the positively skewed stable distribution with stability parameter $\beta = 0.8$, scaled with $n^{-1/\beta}$ to ensure convergence to a stable subordinator. This defines a renewal process, at whose renewal times we assume n i.i.d. magnitudes are drawn from a Generalized Extreme Value Distribution with shape parameter $\xi = 0.7$.

Figure 6.1: A simulated bursty process



Recall that we have defined the exceedance duration of level $\ell \in [x_0, x_F]$ as the random variable

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

and the exceedance as

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

As these random variables are the quantities of interest it is natural for us to then consider the (simulated) sample exceedances and their corresponding durations over the threshold level ℓ.

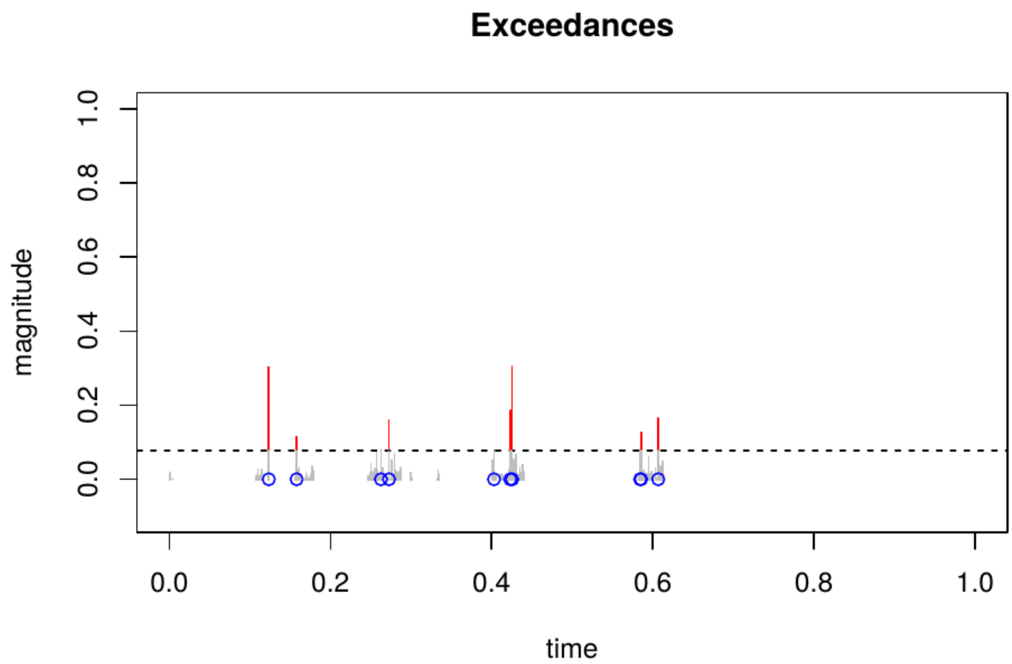


Figure 6.2: Exceedance durations (blue circles) and Exceedance sizes (red lines).

Assume now a time series of magnitudes, and that interest lies in the estimation of the timings of the large magnitudes. Consider a minimum threshold ℓ_0, e.g. at the 95% quantile. Vary the threshold ℓ on the interval $[\ell_0, x_F]$, and consider the resulting sequences of exceedance sizes and exceedance times $\{(X_{\ell,i}, T_{\ell,i})\}$. Due to the renewal property each sequence is i.i.d.. Now $T_{\ell,1}, T_{\ell,2}, \ldots$ can be modelled by a Mittag-Leffler distribution and $J_{\ell,1}, J_{\ell,2}, \ldots$ can be modelled by a Generalised Pareto distribution as shown in chapter **??**.

6.2 Model Fit of Real Data

Chapter 7

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.

[BR77] A. A. Balkema and S. I. Resnick. Max-infinite divisibility. *J. Appl. Probability*, 14(2):309–319, 1977.

[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.

[DR84] Richard Davis and Sidney Resnick. Tail estimates motivated by extreme value theory. *Ann. Statist.*, 12(4):1467–1487, 1984.

[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.

[JW94] Aleksander Janicki and Aleksander Weron. Can one see α-stable variables and processes? *Statist. Sci.*, 9(1):109–126, 1994.

[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.

[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.

[MS12] Mark M. Meerschaert and Alla Sikorskii. *Stochastic models for fractional calculus*, volume 43 of *de Gruyter Studies in Mathematics.* Walter de Gruyter & Co., Berlin, 2012.

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

[Nol15] J. P. Nolan. *Stable Distributions - Models for Heavy Tailed Data.* Birkhauser, Boston, 2015. In progress, Chapter 1 online at academic2.american.edu/∼jpnolan.

[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.

[Res74] Sidney I. Resnick. Inverses of extremal processes. *Advances in Appl. Probability*, 6:392–406, 1974.

[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.

[ST94] Gennady Samorodnitsky and Murad S. Taqqu. *Stable non-Gaussian random processes.* Stochastic Modeling. Chapman & Hall, New York, 1994. Stochastic models with infinite variance.

[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.