THREE PROBLEMS IN QUANTITATIVE RISK MANAGEMENT

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Abhimanyu Mitra
May 2011
This thesis deals with the approximation of the probability of remote risk regions. The simplest example is to compute $P[X > x]$ for a one-dimensional random variable $X$ and a large threshold $x$. Such probabilities give useful measures of risk. We consider three problems related to the approximation of the probability of a risk region.

The first, an important problem in finance and insurance, is to approximate the probability that a sum of losses, $X + Y$, exceeds a large threshold. We investigate a common case where the distribution of $(X, Y)$ belongs to the maximal domain of attraction of a bivariate Gumbel distribution with $X$ and $Y$ being asymptotically independent [18, pages 18, 229] so that both $X$ and $Y$ are in the maximal domain of attraction of the Gumbel distribution. We obtain sufficient conditions to guarantee tail equivalence of $X + Y$ and $X$, that is

$$\lim_{x \to \infty} \frac{P(X + Y > x)}{P(X > x)} \in (0, \infty).$$

Under the further assumption of non-negativity of losses, the result is extended to aggregation of any finite number of losses. We explore the asymptotics of finite linear combinations of losses $\sum_{i=1}^{n} a_i X_i$ with $a_i \geq 0$, $i = 1, 2, \cdots, n$, which we then use to suggest an approximate solution for an optimization problem applicable to portfolio design.

As opposed to aggregation of a fixed number of losses dealt with in the first problem, in the second problem we deal with aggregation of a random number of losses. This problem arises from warranty claims modeling. Consider a retail company, for example a car company, that sells items each of which is covered...
by a warranty for a period $W$. To decide on a reserve for the next quarter, the company has to estimate the quantiles of the distribution of the total warranty cost for the next quarter, based on historical data. Here, each warranty claim arriving in the next quarter is a loss to the retail company and the total cost is the aggregation of such losses. However, the number of claims that will arrive in the next quarter is random. We approximate the distribution of total warranty cost using minimal assumptions on the sales process and the nature of arrival of claims thus making the approximation robust against model error. We suggest a method of computing quantiles of the distribution of the total warranty cost in the next quarter using historical data, which is applied to warranty claims data from a car manufacturer for a single car model and model year.

The third problem deals with joint tail probability estimation, for example $P[Z^1 > x, Z^2 > y]$ for two large thresholds $x$ and $y$. The joint tail probability $P[Z^1 > x, Z^2 > y]$ is a useful measure of risk which helps us understand the tail-dependence of $Z^1$ and $Z^2$. Under the standard model for heavy-tailed losses, multivariate regular variation (abbreviated MRV) [47, page 172] often estimates $P[Z^1 > x, Z^2 > y]$ as zero but hidden regular variation (HRV) [46] offers a refinement of MRV which provides a non-zero and more accurate estimate of $P[Z^1 > x, Z^2 > y]$. In prior work, HRV was defined only on the cone $E^{(2)} = \{x \in [0, \infty]^d : x^{(2)} > 0\}$, where $x^{(2)}$ is the second largest component of $x$. We extend HRV on other sub-cones $E^{(l)} = \{x \in [0, \infty]^d : x^{(l)} > 0\}$ of $E^{(2)}$ as well, $3 \leq l \leq d$, where $x^{(l)}$ is the $l$-th largest component of $x$. For $d > 2$, this extended model of HRV significantly improves the accuracy of the estimates of joint tail probabilities compared to the earlier model of HRV. We suggest some exploratory methods of detecting the presence of HRV on $E^{(l)}$, $2 \leq l \leq d$. Using HRV, we devise a method of estimating joint tail prob-
abilities $P[Z_i^l > x_{i_1}, Z_i^l > x_{i_2}, \ldots, Z_i^l > x_{i_l}]$ for $2 \leq l \leq d$, $1 \leq i_1 < i_2 < \cdots < i_l \leq d$ from data. We apply our method to Internet traffic data to compute a measure of burstiness.
BIOGRAPHICAL SKETCH

Abhimanyu Mitra was born in October, 1982 in Kolkata, India. After completing schooling from Howrah Zilla School, he joined the Bachelor of Statistics (B.Stat.) program at the Indian Statistical Institute (ISI) in July, 2001. Upon graduating in 2004 with a Bachelor’s degree, he decided to continue studying at the same institute for a Master of Statistics (M.Stat.) degree.

After spending five wonderful years at ISI, Abhimanyu joined the School of Operations Research and Information Engineering (ORIE) of Cornell University in August, 2006 to pursue his Ph.D. with concentration in applied probability and statistics.

Abhimanyu has accepted a position of associate in Goldman Sachs and he plans to join there soon after completing his Ph.D.
To my parents
ACKNOWLEDGEMENTS

In my limited experience, research seems like sailing into a vast ocean in search of an unknown land. A sailor should first thank his guiding compass upon arrival to destination and so, I begin my acknowledgement by thanking my advisor Professor Sidney I. Resnick. Without his constant encouragement and guidance, I would have easily lost direction in research and this thesis would have never been completed. I would hardly ever have enough words to thank him.

Professor Michael J. Todd and Professor Robert A. Jarrow deserve special thanks for agreeing to be in my special committee. I have learnt a lot from attending their classes. The discussions of my research with them had always been exciting and in many occasions, filled me with new insights.

Of course, to make a voyage, one needs preparations. I consider myself privileged that throughout my academic career, I have been part of exceptional institutions like the Indian Statistical Institute and the Cornell University. Ex-pounding on the quality of education I received from the faculty of these institutions, would, on one hand, be impossible to fit in these pages, and on the other hand, be a brazen attempt on my part to pretend to do something I know I cannot. So, I choose only to thank all my professors for preparing me well.

The role of my friends and family in my journey go far beyond than what could be reciprocated by a mere thanks. As is characteristic of a long journey, my journey was also sprinkled with a medley of both buoyant and dark hours and my friends and family were always there bootstrapping me out of my melancholy moods. With my obvious weakness in language, making a list of my family members and very good friends and trying to explain what their role had been in my journey, would end in a ludicrous failure and so I refrain from such
an endeavor.

When does a voyage really start? Perhaps, not exactly the time when the ship leaves the shore. I believe, it all begins much earlier with a dream of the voyager to explore. The architects of this dream could perhaps claim all the credit for every success of the voyage and it is certainly not in the hands of the voyager to determine how much contribution they have in the completion of the voyage, let alone thank them.

As the voyager in my story, I dedicate this thesis to the architects of my dream: my parents.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biographical Sketch</td>
<td>iii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>v</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xi</td>
</tr>
<tr>
<td><strong>1 Overview</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Aggregation of rapidly varying risks and asymptotic independence</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Modeling total expenditure on warranty claims</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Hidden regular variation: detection and estimation</td>
<td>4</td>
</tr>
<tr>
<td><strong>2 Aggregation of rapidly varying risks and asymptotic independence</strong></td>
<td>6</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>6</td>
</tr>
<tr>
<td>2.2 Asymptotic tail probability for aggregated risk</td>
<td>9</td>
</tr>
<tr>
<td>2.2.1 Asymptotic tail probability for the sum of two random variables</td>
<td>9</td>
</tr>
<tr>
<td>2.2.2 Asymptotic tail probability for the sum of more than two non-negative random variables</td>
<td>20</td>
</tr>
<tr>
<td>2.3 Examples</td>
<td>24</td>
</tr>
<tr>
<td>2.4 Linear combinations of random variables with non-negative coefficients</td>
<td>31</td>
</tr>
<tr>
<td>2.4.1 Tail equivalent cases</td>
<td>31</td>
</tr>
<tr>
<td>2.4.2 One special case where the distributions are possibly not tail equivalent</td>
<td>39</td>
</tr>
<tr>
<td>2.5 An optimization problem</td>
<td>42</td>
</tr>
<tr>
<td>2.5.1 An approximate solution</td>
<td>43</td>
</tr>
<tr>
<td>2.5.2 A special case</td>
<td>45</td>
</tr>
<tr>
<td>2.6 Simulation studies</td>
<td>45</td>
</tr>
<tr>
<td>2.6.1 Where is the approximation good?</td>
<td>46</td>
</tr>
<tr>
<td>2.6.2 How good is the portfolio suggestion?</td>
<td>48</td>
</tr>
<tr>
<td>2.7 Concluding Remarks</td>
<td>53</td>
</tr>
<tr>
<td><strong>3 Modeling total expenditure on warranty claims</strong></td>
<td>56</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>56</td>
</tr>
<tr>
<td>3.1.1 Outline</td>
<td>58</td>
</tr>
<tr>
<td>3.1.2 Notation</td>
<td>58</td>
</tr>
<tr>
<td>3.2 Non-renewable free replacement warranty policy</td>
<td>59</td>
</tr>
<tr>
<td>3.2.1 Assumptions</td>
<td>60</td>
</tr>
<tr>
<td>3.2.2 Approximation of the distribution of total cost on warranty claims</td>
<td>62</td>
</tr>
<tr>
<td>3.3 Non-renewable pro-rata warranty policy</td>
<td>66</td>
</tr>
</tbody>
</table>
3.3.1 Approximation of the distribution of total cost on warranty claims ........................................ 67
3.4 Examples of sales processes consistent with our assumptions ...................................................... 68
3.5 Estimation procedure ....................................................................................................................... 70
3.5.1 Free replacement policy .............................................................................................................. 71
3.5.2 Pro-rata policy ........................................................................................................................ 78
3.6 Computational example .................................................................................................................. 79
3.6.1 Analysis of the claim size distribution ......................................................................................... 80
3.6.2 Analysis of the distribution of the times of claims measure $M(\cdot)$ ........................................... 83
3.6.3 Analysis of the distribution of the sales process ........................................................................ 85
3.6.4 Estimation of quantiles .............................................................................................................. 86
3.7 Concluding remarks ...................................................................................................................... 90
3.8 Proofs ........................................................................................................................................ 91
3.9 Glossary of notation for Chapter 3 ............................................................................................... 104

4 Hidden regular variation: detection and estimation ........................................................................ 107
4.1 Introduction ................................................................................................................................. 107
4.1.1 Outline .................................................................................................................................. 111
4.1.2 Notation ................................................................................................................................ 112
4.2 Hidden regular variation ............................................................................................................. 115
4.2.1 Hidden regular variation on $E^{(2)}$ ...................................................................................... 115
4.2.2 Hidden regular variation beyond $E^{(2)}$ ................................................................................ 117
4.2.3 Examples ............................................................................................................................... 121
4.3 Exploratory detection and estimation techniques .......................................................................... 124
4.3.1 Decomposition of the limit measure $\nu^{(l)}(\cdot)$ .................................................................. 124
4.3.2 Detection of HRV on $E^{(l)}$ and estimation of $\nu^{(l)}(\cdot)$ ....................................................... 126
4.4 A different representation of the hidden angular measure $S^{(l)}(\cdot)$ ......................................... 132
4.4.1 The transformation ................................................................................................................. 132
4.4.2 Estimation of $\tilde{S}^{(l)}(\cdot)$ ..................................................................................................... 134
4.4.3 Supports of transformed (hidden) angular measure $\tilde{S}^{(l)}(\cdot)$ ........................................... 134
4.4.4 Lines through $\infty$ ................................................................................................................ 137
4.5 Deciding finiteness of $\nu^{(l)}(\{x \in E^{(l)} : ||x|| > 1\})$ ................................................................. 140
4.5.1 A moment condition ................................................................................................................. 140
4.5.2 A particular construction ........................................................................................................ 142
4.6 Computation of probabilities of risk sets ...................................................................................... 145
4.6.1 At least one risk is large .......................................................................................................... 145
4.6.2 Linear combination of risks ................................................................................................... 146
4.7 Computational examples .............................................................................................................. 150
4.7.1 Simulated data ......................................................................................................................... 150
4.7.2 Internet traffic data ................................................................................................................ 154
4.8 Concluding remarks ..................................................................................................................... 156
4.9 Proofs ........................................................................................................................................ 160
Some convergence concepts and distributional relations 168
A.1 Tail equivalence .............................. 168
A.2 Subexponential distributions ................ 168
A.3 Vague convergence .......................... 169

Maximal domain of attraction and regular variation 170
B.1 Maximal domain of attraction ............... 170
B.2 Regular variation ............................. 171
LIST OF TABLES

2.1 Simulation results to judge goodness of approximation when \( p = -1 \) .......................................................... 48
2.2 Simulation results to judge goodness of approximation when \( p = -0.9 \) ............................................................ 49
2.3 Simulation results to judge goodness of approximation when \( p = 0 \) ................................................................. 50
2.4 Simulation results to judge goodness of approximation when \( p = 0.9 \) ................................................................. 51
2.5 Simulation study to judge effectiveness of approximate optimization when \( p = -0.9 \) ............................................. 53
2.6 Simulation study to judge effectiveness of approximate optimization when \( p = 0 \) ..................................................... 54
2.7 Simulation study to judge effectiveness of approximate optimization when \( p = 0.9 \) ..................................................... 54
3.1 Summary statistics for claim size data ............................................. 81
3.2 Summary statistics for claim size data of size less than 500 ........... 82
3.3 Estimated parameters \( c_1, c_2, \hat{\mu}, \hat{\sigma}^2 \) ................................................. 87
3.4 Quantiles for the total cost on warranty claims .......................... 89
3.5 Values of \( \{ A_i F_p(COST^n(P)), i = 1, 2, P = [0, T], [T, 2T] \} \) ..... 90
# LIST OF FIGURES

3.1 Density and QQ plot of the claim size distribution (k = 5000, n = 49323), Left - Density plot, Right - QQ plot. 81
3.2 Density and QQ plot of the distribution of claim sizes which are less than 500 (k = 5000, n = 48864), Left - Density plot, Right - QQ plot. 82
3.3 Plot of \((i, \hat{m}_1((i - 1, i))\): Left - \(\{i = 0, 1, \cdots, 1096\}\) and Right - \(\{i = 1, \cdots, 1095\}\). 85
3.4 Observed sales process with fitted Bass: Left - Daily counts, Right - 12-day counts. 86
3.5 Time and QQ plot of \(\{j_t\}\): Left - Time plot, Right - QQ plot. 86
4.1 The set \(\delta N^{(2)}\) (shaded region), when \(d = 3\). 111
4.2 Hill plot of \(\alpha^{(2)}\) and estimated and actual transformed hidden angular densities. 151
4.3 Plots of estimates of \(P[X_1 > 100, Y_1 > \sqrt{10}]\) for different values of \(k\) (sample size = 5000) using both our and H-R(Heffernan-Resnick [27]) estimator. 153
4.4 Plots of estimates of \(P[X_1 > 100, Y_1 > \sqrt{10}]\) for different values of \(k\) (sample size = 500) using both our and H-R(Heffernan-Resnick [27]) estimator. 154
4.5 Hill plot of \(\alpha^{(2)}\) and estimated transformed hidden angular densities. 156
4.6 Plots of estimates of \(P[\text{Size} > 2 \times 10^7, \text{Rate} > 10^5]\) for different values of \(k\) using our estimator and H-R(Heffernan-Resnick [27]) estimator. 157
4.7 Plots of estimates of \(P[\text{Size} > 2 \times 10^8, \text{Rate} > 10^6]\) for different values of \(k\) using our estimator and H-R(Heffernan-Resnick [27]) estimator. 157
Risk management is central to business, finance, insurance [40, Chapter 1] and many engineering disciplines concerned with structural or environmental risk like hydrology [17] and environmental protection [57]. Risk management requires accurate evaluation of risks and often leads to estimation of probabilities of rare events such as tail probabilities for high thresholds. An example in one-dimensional estimation would be to estimate \( P[X > x] \) for large thresholds \( x \). This thesis concentrates on approximation of tail probabilities with high thresholds. Such probabilities give useful measures of risk. We discuss three problems related to tail probability estimation. With some abuse of terminology (though consistent with existing literature), we call the random vectors denoting losses etc. as risk vectors and also by risk we mean some measure of risk computed using the distribution of the random (risk) vectors, for example Value-at-Risk (VaR), expected shortfall or tail probability.

1.1 Aggregation of rapidly varying risks and asymptotic independence

In finance and insurance, it is important to estimate the probability that a sum of risks or losses, \( X + Y \), exceeds a large threshold. This has been considered under various assumptions by many researchers; see for example [1, 2, 33].

Asymptotic analysis requires the robust assumption that the joint distribution \( H(\cdot) \) of \( (X, Y) \) belongs to some maximal domain of attraction, that is for all
continuity points \((x_1, x_2)\) of a bivariate distribution \(G(\cdot)\),

\[
\lim_{n \to \infty} H^\alpha(a_n^1 x_1 + b_n^1, a_n^2 x_2 + b_n^2) = G(x_1, x_2)
\]  

(1.1.1)

and both the marginal distributions of \(G(\cdot), G_1(\cdot)\) and \(G_2(\cdot)\), are non-degenerate extreme value distributions [18, page 208]. Researchers have divided the huge class of distributions satisfying (1.1.1) into different subclasses and approximated the probability of \(X + Y\) exceeding a fixed large threshold; a survey can be found in [1]. But, in one important subclass \(\mathcal{A}\) of risk distributions, where (1.1.1) is satisfied with \(G(x_1, x_2) = G_1(x_1)G_2(x_2)\), \((x_1, x_2) \in \mathbb{R}^2\) and \(G_1(x) = G_2(x) = \exp(-e^{-x}), x \in \mathbb{R}\) (Gumbel distribution), no such result was known. If the distribution of \((X, Y)\) belongs to \(\mathcal{A}\), then both the risks \(X\) and \(Y\) are rapidly varying or \(-\infty\)-varying [45, page 53] and \(X\) and \(Y\) are asymptotically independent [18, page 229].

We give a set of conditions on the joint distribution of \((X, Y)\), which guarantees tail equivalence of \(X + Y\) and \(X\), that is

\[
\lim_{x \to \infty} \frac{P(X + Y > x)}{P(X > x)} = 1 + c,
\]

assuming \(c = \lim_{x \to \infty} P(Y > x)/P(X > x)\). Establishing tail equivalence of \(X + Y\) and \(X\) is fundamental in designing good algorithms for simulating rare events such as \(P(X + Y > x)\) for large thresholds \(x\). Under the further assumption of non-negativity of risks, the result is extended to the aggregation of any finite number of risks. We show that if \(X_1, X_2, \cdots, X_n\) are pairwise tail equivalent and satisfy our set of conditions, then the finite linear combination of losses \(\sum_{i=1}^n a_i X_i\) with \(a_i \geq 0, \ i = 1, 2, \cdots, n\), is tail equivalent to \(a_k X_k\), where \(k = \arg \max_i \{a_i\}\). Using this result, we suggest an approximate solution for an optimization problem applicable to portfolio design.
1.2 Modeling total expenditure on warranty claims

A retail company selling items with warranties is exposed to the risk of future warranty claims. The company makes a reserve to cover the warranty claims in the next quarter, say \([0, T]\), and must estimate the total cost on such claims based on historical data; see [35]. Here also the total expenditure of the company on warranty claims in \([0, T]\) is the aggregation of the expenditures on individual claims, only with the added complexity that the the number of claims in \([0, T]\) is random.

Naturally, the company wishes to know the reserve level \(R\) such that the probability that the total warranty cost \(COST\) in \([0, T]\) exceeds \(R\), that is \(P(COST > R)\), is very small. In other words, the company wants to know a very high quantile of the distribution of \(COST\), a natural measure of risk for the company. We may also view the very high quantiles of the distribution of \(COST\) as a measure of risk to the company.

We approximate the distribution of \(COST\). Compared to previous work [35], our modeling requires minor assumptions on the sales process and the nature of arrival of claims, thus making our approximation robust against model error. We consider two kinds of warranty policies, the non-renewing free replacement warranty policy and the non-renewing pro-rata warranty policy. In each case, we approximate the distribution of \(COST\) by a normal or stable distribution. We suggest a method of estimating the parameters of the approximating normal or stable distribution from historical data, after which computation of quantiles is a routine procedure. Our method of quantile computation is applied to warranty claims data from a car manufacturer for a single car model and model year.
1.3 Hidden regular variation: detection and estimation

In risk management, it is important to accurately calculate risk and when a model fails to capture some hidden risks, we seek a refinement of the model. For some risk vector $\mathbf{Z} = (Z^1, Z^2)$ and two large thresholds $x$ and $y$, the joint tail probability $P[Z^1 > x, Z^2 > y]$ is a measure of risk which reflects the tail-dependence of $Z^1$ and $Z^2$. Hidden regular variation (HRV) [46] is a refinement of multivariate regular variation (MRV) [47, page 172], which provides more accurate estimates of $P[Z^1 > x, Z^2 > y]$ than MRV for large thresholds $x$ and $y$.

To introduce MRV and HRV, let us briefly define regular variation on cones. Let $\mathbb{C}$ be a cone in $[0, \infty)^d$, meaning $x \in \mathbb{C}$ implies $tx \in \mathbb{C}$ for $t > 0$. Denote the set of all Radon measures on $\mathbb{C}$ by $M_+(\mathbb{C})$. The distribution of a random vector $\mathbf{Z} = (Z^1, Z^2, \cdots, Z^d)$ is regularly varying on $\mathbb{C}$ if there exist a scaling function $g(t) \uparrow \infty$, and a non-zero Radon measure $\chi(\cdot) \in M_+(\mathbb{C})$ such that

$$
tP \left[ \frac{\mathbf{Z}}{g(t)} \in \cdot \right] \stackrel{\nu}{\rightarrow} \chi(\cdot)$$

in $M_+(\mathbb{C})$, where $\rightarrow$ denotes vague convergence [48].

MRV is a standard way to model non-negative heavy-tailed risks. The distribution of a $d$-dimensional risk vector $\mathbf{Z} = (Z^1, Z^2, \cdots, Z^d)$ has MRV if it is regularly varying on the cone $\mathbb{E} = [0, \infty)^d \setminus \{(0, 0, \cdots, 0)\}$ as in (1.3.1) with limit measure $\nu(\cdot)$. However, it may happen that $\nu(\cdot)$ puts zero mass on a sub-cone of $\mathbb{E}$. HRV allows modeling a different regular variation on a sub-cone of $\mathbb{E}$, where $\nu(\cdot)$ puts zero mass. Thus, HRV helps in detecting some finer structures ignored by MRV; see [27]. The presence of HRV is found in data from the international stock market [43], Internet traffic [28] and coastal flooding [10] among others.
In prior work, HRV was defined on the sub-cone $\mathbb{E}^{(2)} = \{x \in \mathbb{E} : x^{(2)} > 0\}$, where $x^{(2)}$ is the second largest component of $x$, which forms a subfamily of models possessing MRV and asymptotic independence [47, pages 323-325]. We extend models of HRV on other sub-cones $\mathbb{E}^{(l)} = \{x \in \mathbb{E} : x^{(l)} > 0\}$ of $\mathbb{E}$ as well, where $x^{(l)}$ is the $l$-th largest component of $x$, $3 \leq l \leq d$, and show with an example that asymptotic independence is not a necessary condition for HRV on $\mathbb{E}^{(l)}$, $3 \leq l \leq d$. For $d > 2$, this extended model of HRV significantly improves the accuracy of the estimates of joint tail probabilities compared to the earlier model of HRV. We suggest some exploratory methods of detecting the presence of HRV on $\mathbb{E}^{(l)}$, $2 \leq l \leq d$. In the presence of HRV on $\mathbb{E}^{(l)}$, $2 \leq l \leq d$, we devise a method of estimating joint tail probabilities $P[Z^{i_1} > x_{i_1}, Z^{i_2} > x_{i_2}, \ldots, Z^{i_l} > x_{i_l}]$ for $2 \leq l \leq d$, $1 \leq i_1 < i_2 < \cdots < i_l \leq d$, from data and show that it indeed provides better estimates than those provided by MRV. We apply our method to Internet traffic data to compute the probability that both the size of the file transferred and the rate of transfer are very high, leading to a measure of burstiness.
CHAPTER 2
AGGREGATION OF RAPIDLY VARYING RISKS AND ASYMPTOTIC INDEPENDENCE

2.1 Introduction

Estimating the probability that a sum of risks \( X + Y \) exceeds a large threshold is important in finance and insurance, and hence much applied probability research has been dedicated to this goal. Recent results are found in [1, 2, 5, 24, 33, 34, 59]. Approximating this probability helps us evaluate risk measures for investment portfolios as well as estimating credit risk.

The problem is reasonably well understood when risks have regularly varying marginal distributions but another important large class of risk distributions is the maximal domain of attraction of the Gumbel distribution, denoted \( MDA(\Lambda) \), where

\[
\Lambda(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R},
\]

and \( MDA(\Lambda) \) is the class of distributions \( F(\cdot) \) for which there exist \( a_n > 0, b_n \in \mathbb{R} \) such that

\[
\lim_{n \to \infty} n(1 - F(a_n x + b_n)) = \lim_{n \to \infty} n\bar{F}(a_n x + b_n) = e^{-x}, \quad x \in \mathbb{R} \quad (2.1.1)
\]

[45, page 38]. If \( X \sim F(\cdot) \) and \( F(\cdot) \in MDA(\Lambda) \), we also write \( X \in MDA(\Lambda) \). It is also well known that the risks having distribution in \( MDA(\Lambda) \) are rapidly varying, that is \(-\infty\)-varying [45, page 53]. Within the class of risks \((X, Y)\) with marginal distributions \( F_1(\cdot), F_2(\cdot) \in MDA(\Lambda) \), results on aggregation of risks are known when \( X \) and \( Y \) are independent. However, actual risks are often not independent and a somewhat weaker concept called asymptotic independence allows risks to
be modeled as dependent and is more practical in many modeling situations. Risks $X$ and $Y$ in a maximal domain of attraction are \textit{asymptotically independent} if (1.1.1) is satisfied with $G(x_1, x_2) = G_1(x_1)G_2(x_2)$ for $(x_1, x_2) \in \mathbb{R}^2$, where $H(\cdot)$ is the joint distribution of $X$ and $Y$ and both $G_1(\cdot)$ and $G_2(\cdot)$ are non-degenerate extreme value distributions [18, page 229]. There are also results on the aggregation of risks in the absence of asymptotic independence where the analogue of (1.1.1) holds but with a limit distribution which is not a product; see [33].

We consider the case where the risks $X, Y$ are asymptotically independent with marginal distributions $F_1(\cdot), F_2(\cdot) \in MDA(\Lambda)$. We also allow one marginal tail to be lighter and the distribution with lighter tail does not necessarily belong to the maximal domain of attraction of the Gumbel distribution.

Within the class of vectors $(X, Y)$ satisfying asymptotic independence and marginal distributions $F_1(\cdot), F_2(\cdot) \in MDA(\Lambda)$, two prominent but very distinct behaviors have been observed.

1. First, suppose $(X, Y)$ are two i.i.d. risks with common distribution $F_1(\cdot)$ which is subexponential [23, page 39] and $F_1(\cdot) \in MDA(\Lambda)$. Then $X$ and $Y$ are certainly asymptotically independent and

$$
\lim_{x \to \infty} \frac{P(X + Y > x)}{P(X > x)} = 2.
$$

(2.1.2)

So one possible behavior is that the sum has a distribution which is tail equivalent to the distribution of a summand.

2. Very different tail behavior is exhibited in Theorem 2.10 of [1], where the authors exhibit a distribution of $(X, Y)$, with $X$ and $Y$ being asymptotically independent and identically distributed with common distribution $F_1(\cdot) \in$
In Section 2, we give a set of conditions on the joint distribution of \((X, Y)\), guaranteeing behavior of the first sort, namely,

\[
\lim_{x \to \infty} \frac{P(X + Y > x)}{P(X > x)} = 1 + c,
\]

where \(c = \lim_{x \to \infty} P(Y > x)/P(X > x)\), the limit being assumed to exist. If \(c \in (0, \infty)\), our conditions imply that \(X, Y\) are asymptotically independent and each belongs to the maximal domain of attraction of the Gumbel. When \(X, Y\) are identically distributed, (2.1.2) holds. Under the further assumption of non-negativity of risks, the result is extended for the case of more than two risks.

In Section 3, we provide examples of distributions which satisfy our conditions. The examples include cases where the marginal distributions of \(X\) and \(Y\) are subexponential and also cases where they are not. We also show one example which does not satisfy our conditions but yet exhibits the tail equivalence between the distribution of the sum and that of the summand. Thus, our conditions are only sufficient. In Section 4, we summarize asymptotic behavior of finite linear combinations of risks with non-negative coefficients. In Section 5, we suggest approximate solutions for an optimization problem which is related to portfolio design. The chapter closes with concluding remarks and a brief summary of numerical experiments which give a feel for whether asymptotic equivalence is a suitable numerical approximation for exceedance probabilities of aggregated risks.
2.2 Asymptotic tail probability for aggregated risk

2.2.1 Asymptotic tail probability for the sum of two random variables

We give conditions guaranteeing (2.1.3), where the constant $c$ satisfies only $c = \lim_{x \to \infty} P(Y > x)/P(X > x) \in [0, \infty)$. When $c \in (0, \infty)$, $X$ and $Y$ are called tail equivalent [50] and then our conditions guarantee that both the marginal distributions $F_1(\cdot), F_2(\cdot) \in MDA(\Lambda)$ and $X$ and $Y$ are asymptotically independent. When $c = 0$, our result extends to the case where $F_2(\cdot)$, the marginal distribution of $Y$, does not belong to the maximal domain of attraction of the Gumbel distribution and where $X$ and $Y$ need not be asymptotically independent.

Assumptions.

Suppose $(X, Y)$ is a pair of random variables satisfying the following set of assumptions.

1. The random variable $X$ has a distribution $F_1(\cdot)$ whose right endpoint $x_0$ is infinite; that is,

$$x_0 = \sup\{x : F_1(x) < 1\} = \infty. \tag{2.2.1}$$

Further $F_1(\cdot) \in MDA(\Lambda)$ so that (2.1.1) is satisfied with centering constants $b_n \in \mathbb{R}$ and scaling constants $a_n > 0$. Equivalently ([15], [45, page 28, 40-43]) there exists a self-neglecting auxiliary function $f(\cdot)$ with its derivative

9
converging to 0, such that
\[
\lim_{t \to \infty} \frac{\bar{F}_1(t + xf(t))}{\bar{F}_1(t)} = e^{-x}.
\] (2.2.2)

2. The random variables \(X\) and \(Y\) have distribution functions \(F_1(\cdot)\) and \(F_2(\cdot)\) such that
\[
\lim_{x \to \infty} \frac{\bar{F}_2(x)}{\bar{F}_1(x)} = c \in [0, \infty).
\]

3. The conditional distribution of \(Y\) given \(X > x\), satisfies for all \(t > 0\),
\[
\lim_{x \to \infty} P(|Y| > tf(x)|X > x) = 0,
\]
where \(f(x)\) is the auxiliary function corresponding to the distribution of \(X\) given in (2.2.2),

4. Symmetrically assume for all \(t > 0\),
\[
\lim_{x \to \infty} P(|X| > tf(x)|Y > x) = 0.
\]

5. For some \(L > 0\), suppose
\[
\lim_{x \to \infty} \frac{P(Y > Lf(x), X > Lf(x))}{P(X > x)} = 0.
\]

The main result.

The assumptions allow us to conclude aggregated risks are essentially tail equivalent to individual risks.

**Theorem 2.2.1.** Under the Assumptions in Section 2.2.1,
\[
\lim_{x \to \infty} \frac{P(X + Y > x)}{P(X > x)} = (1 + c).
\] (2.2.3)
Comments on the assumptions.

Before giving a proof of Theorem 2.2.1, we discuss implications of the assumptions.

Remark 2.2.2. 1. When $F_1(\cdot) \in MDA(\Lambda)$, we may choose $a_n, b_n$ appearing in (2.1.1) as $b_n = b_{F_1}(n) = \inf \left\{ s : F_1(s) \geq 1 - \frac{1}{n} \right\}$, $a_n = f(b_n)$, where $f(\cdot)$ is defined in (2.2.2); see [45, page 40] or [18]. So, from (2.2.1) and the definition of $b_n$, it follows that $\lim_{n \to \infty} b_n = \infty$. Also, since $\lim_{x \to \infty} f(x)/x = 0$ [45, page 40], $\lim_{n \to \infty} a_n/b_n = 0$.

2. If $c \in (0, \infty)$, then our assumptions guarantee both marginal distributions $F_1(\cdot), F_2(\cdot) \in MDA(\Lambda)$ and also that $(X, Y)$ are asymptotically independent. From Assumption 1, $F_1(\cdot) \in MDA(\Lambda)$ and since $F_1(\cdot)$ and $F_2(\cdot)$ are tail equivalent, from [50] we get that $F_2(\cdot) \in MDA(\Lambda)$. For asymptotic independence, define

$$b_{F_1}(t) = \inf \left\{ s : \frac{1}{1-F_1}(s) \geq t \right\} = \left( \frac{1}{1-F_1} \right)^-(t) \quad (2.2.4)$$

and similarly $b_{F_2}(t)$. From [18, page 229], if $F_1(\cdot), F_2(\cdot) \in MDA(\Lambda)$ and

$$\lim_{t \to \infty} \frac{P(X > b_{F_1}(t), Y > b_{F_2}(t))}{P(X > b_{F_1}(t))} = 0, \quad (2.2.5)$$

then $(X, Y)$ are asymptotically independent. When $c \in (0, \infty)$, Assumption 3 implies (2.2.5). To verify this, note first that Assumption 3 implies

$$\lim_{x \to \infty} \frac{P(X > x, Y > x)}{P(X > x)} \leq \lim_{x \to \infty} \frac{P(X > f(x), Y > x)}{P(X > x)} = 0, \quad (2.2.6)$$

since $f(x)/x \to 0$ as $x \to \infty$ [45, page 40]. If $c > 1$, then for sufficiently large $t$, $b_{F_1}(t) \leq b_{F_2}(t)$ and therefore, using (2.2.6),

$$\lim_{t \to \infty} \frac{P(X > b_{F_1}(t), Y > b_{F_2}(t))}{P(X > b_{F_1}(t))} \leq \lim_{t \to \infty} \frac{P(X > b_{F_1}(t), Y > b_{F_2}(t))}{P(X > b_{F_1}(t))}$$
\[
= \lim_{t \to \infty} \frac{P(X > t, Y > t)}{P(X > t)} = 0,
\]
as required. A similar verification can be constructed for the case \(0 < c < 1\).

For \(c = 1\), \(b_{F_1}(t) \sim b_{F_2}(t)\). Hence,

\[
\frac{f(b_{F_1}(t))}{b_{F_2}(t)} \sim \frac{f(b_{F_1}(t))}{b_{F_1}(t)} \to 0.
\]

So,

\[
\lim_{t \to \infty} \frac{P(X > b_{F_1}(t), Y > b_{F_2}(t))}{P(X > b_{F_1}(t))} \leq \lim_{t \to \infty} \frac{P(X > b_{F_1}(t), Y > f(b_{F_1}(t)))}{P(X > b_{F_1}(t))} = 0 \text{ (by Assumption 3 and (2.2.1)).}
\]

3. The auxiliary function \(f(\cdot)\) can be replaced by any asymptotically equivalent function \(\tilde{f}(\cdot)\); that is, if \(\lim_{x \to \infty} \tilde{f}(x)/f(x) = 1\), and if Assumptions 3, 4, 5 hold with \(f(\cdot)\), they also hold with \(\tilde{f}(\cdot)\) replacing \(f(\cdot)\) and vice versa. Since the mean excess function

\[
e(x) = E(X - x | X > x)
\]
is asymptotically equivalent to any auxiliary function \(f(x)\) ([23, page 143], [45, page 48]), \(e(x)\) can also be taken as an auxiliary function.

4. If \(c = \lim_{x \to \infty} \tilde{F}_2(x)/\tilde{F}_1(x) = 0\), we do not need Assumption 4 to conclude our result.

5. An easier proof of the result can be given if Assumption 5 holds for all \(L > 0\). But here we provide an example to show the importance of the weak version of Assumption 5.

**Example 2.2.3.** We define two random variables \(X\) and \(Y\) as

\[
X = -\log(U), \quad Y = -\log(1 - U), \quad \text{where } U \sim \text{Uniform}(0, 1).
\]
In this case both $X$ and $Y$ have the same distribution, which is Exponential(1). So, the auxiliary function is $f(x) = (1 - F_1(x))/F'_1(x) = 1$ [45, page 40]. Choose $L$ such that $\exp(-L) = \frac{3}{4}$, and
\[
\frac{P(X > Lf(x), Y > Lf(x))}{P(X > x)} = \frac{P(U < \exp(-L), 1 - U < \exp(-L))}{P(X > x)} = \frac{P(\frac{1}{4} < U < \frac{3}{4})}{2P(X > x)} \to \infty.
\]
Therefore, this particular choice of $L$ does not satisfy Assumption 5. The distribution of $(X, Y)$ is a special case of Example 2.3.4 which discusses certain $L$ which do satisfy assumption 5.

6. If, however, both $X$ and $Y$ are non-negative risks, and Assumption 5 is strengthened to hold for all $L > 0$, then Assumptions 3 and 4 will be automatically satisfied. The proof of this follows from $\lim_{x \to \infty} f(x)/x = 0$.

7. Similar limit results are found in Lemma 2.7 of [1] and Theorem 2.1 of [34]. They have assumed that one of the marginal distributions of the two asymptotically independent variables $X$ and $Y$, say the distribution of $X$, is subexponential (that is $X \in S$, where $S$ is the set of all subexponential distributions [23, page 39]), and worked on finding conditions for the tail equivalence of the marginal distribution of $X$ and the sum $X + Y$. Our assumptions are different: we assume that one of the marginal distributions of the two asymptotically independent variables $X$ and $Y$, say the distribution of $X$, belongs to the domain of attraction of Gumbel, that is $X \in MDA(\Lambda)$. We do not assume the marginal distribution of $X$ is subexponential.

In examples where the marginal distributions of the two asymptotically independent and identically distributed random variables $X$ and $Y$ belong to the class $MDA(\Lambda) \cap S$, an issue is the relative strength of our conditions.
versus those of Theorem 2.1 of [34]. We can not show either set of conditions implies the other. However, below we present an example which satisfies our set of conditions, but does not satisfy the set of conditions given in Theorem 2.1 of [34]. Thus our set of conditions is not stronger.

**Example 2.2.4.** Suppose \( X = \exp(X_1), \ Y = \exp(X_2), \) where \((X_1, X_2)\) is bivariate normal with correlation \( \rho \in (0, 1) \). For simplicity, assume each \( X_i \) has mean 0 and variance 1. The lognormal distribution belongs to the class \( MDA(\Lambda) \cap S \) [23, page 39]. In Example 2.3.6, we show \((X, Y)\) satisfy our set of conditions. Here we show that this example does not satisfy Assumption 2.1 of [34], that is for all \( x^* > 0 \),

\[
\limsup_{x \to \infty} \sup_{x^* \leq x \leq t} \frac{P(Y > x - t|X = t)}{P(Y > x - t)} = \infty. \tag{2.2.7}
\]

From the exchangeability of \( X \) and \( Y \), it is obvious that (2.2.7) holds even if the role of \( X \) and \( Y \) is interchanged.

\[
\sup_{x^* \leq x \leq t} \frac{P(Y > x - t|X = t)}{P(Y > x - t)} = \sup_{x^* \leq x \leq t} \frac{\Phi\left(\frac{\log(x - t) - \rho \log t}{\sqrt{1 - \rho^2}}\right)}{\Phi\left(\log(x - t)\right)} \geq \frac{\Phi\left(\frac{\log(x/2) - \rho \log(x/2)}{\sqrt{1 - \rho^2}}\right)}{\Phi\left(\log(x/2))\right)} \Rightarrow \Phi\left(\frac{1 - \rho}{\sqrt{1 - \rho^2}} \log(x/2)\right) \to \infty. \tag{2.2.8}
\]

The inequality above follows from choosing \( x \) large enough so that \( x/2 > x^* \) and putting \( t = x/2 \). The last convergence follows from the fact that the normal distribution belongs to the class \( MDA(\Lambda) \) and hence \( \Phi \) is \(-\infty\)-varying [45, page 53]. Note, \( 0 < \rho < 1 \) entails \( \frac{1 - \rho}{\sqrt{1 - \rho^2}} < 1 \). Hence, from (2.2.8) it is obvious that (2.2.7) holds.
Proof of Theorem 2.2.1.

We prove Theorem 2.2.1 using a Proposition and a Lemma, which we prove first. Note, we do not need the assumption that the marginal distributions are subexponential, which is a necessary condition in the case where $X$ and $Y$ are independent.

**Proposition 2.2.5.** Under Assumptions 1 and 3 of Section 2.2.1, we have

$$
\lim_{n \to \infty} P(Y \leq a_n z | X > a_n x + b_n) = 1_{\{z \geq 0\}}, \quad z \neq 0, \quad x \in \mathbb{R}.
$$

(2.2.9)

and from Assumptions 1 and 4 of Section 2.2.1, we have

$$
\lim_{n \to \infty} P(X \leq a_n z | Y > a_n x + b_n) = 1_{\{z \geq 0\}}, \quad z \neq 0, \quad x \in \mathbb{R}.
$$

(2.2.10)

**Proof.** From Remark 2.2.2 (1), we choose $a_n$ as $a_n = f(b_n)$. Note that for $x \in \mathbb{R},$

$$
\lim_{n \to \infty} (b_n + xa_n) = \lim_{n \to \infty} b_n(1 + xa_n/b_n) = \infty,
$$

since $\lim_{n \to \infty} b_n = \infty, \lim_{n \to \infty} a_n/b_n = 0$ [Remark 2.2.2 (1)]. Now, note that the self-neglecting property of the auxiliary function $f(\cdot)$ defined in (2.2.2), that is

$$
\lim_{t \to \infty} \frac{f(t + xf(t))}{f(t)} = 1, \quad x \in \mathbb{R},
$$

coupled with the fact $\lim_{n \to \infty} b_n = \infty$ [Remark 2.2.2 (1)], implies

$$
\lim_{n \to \infty} \frac{f(b_n + xf(b_n))}{f(b_n)} = 1, \quad x \in \mathbb{R}.
$$

Thus, for $\epsilon \in (0, 1)$, there exists $N$ such that for $n \geq N$, $\frac{f(b_n)}{f(b_n + xf(b_n))} \geq \epsilon$. Assume $z > 0, \epsilon \in (0, 1)$. Using Assumption 3, we get as $n \to \infty,$

$$
P(Y > z\epsilon f(b_n + xf(b_n)) | X > f(b_n)x + b_n)
\leq P(|Y| > z\epsilon f(b_n + xf(b_n)) | X > f(b_n)x + b_n) \to 0.
$$

(2.2.11)
So, for \( n \geq N \),

\[
P(Y \leq a_n z | X > a_n x + b_n) = P(Y \leq f(b_n) z | X > f(b_n) x + b_n)
\]

\[
= P(Y \leq f(b_n + x f(b_n)) \frac{f(b_n)}{f(b_n + x f(b_n))} | X > f(b_n) x + b_n)
\]

\[
\geq P(Y \leq z \varepsilon f(b_n + x f(b_n)) | X > f(b_n) x + b_n)
\]

\[
= 1 - P(Y > z \varepsilon f(b_n + x f(b_n)) | X > f(b_n) x + b_n),
\]

which converges to 1 from (2.2.11). To summarize, as required, we have for \( z > 0 \),

\[
P(Y \leq a_n z | X > a_n x + b_n) \rightarrow 1 = 1_{\{z > 0\}}, \quad n \to \infty.
\]

The argument is similar for \( z < 0 \).

The second part is proved similarly. \(\square\)

**Lemma 2.2.6.** (i) Assumptions 1, 2, and 3 of Section 2.2.1 imply that the sequence of measures

\[
nP[a_n^{-1}(X - b_n, Y) \in (dx, dy)]
\]

converges vaguely [47, page 48] on \(([-M, \infty] \times [-\infty, \infty])\) as \( n \to \infty \), to the limit measure \( m_{1,\infty}(dx, dy) = e^{-y} dx e_0(dy) \), for some \( M > L \) (from Assumption 5 of Section 2.2.1) such that \(-M \) is a continuity point of \( [a_n^{-1}(X - b_n)] \) for all \( n \).

(ii) Assumptions 1, 2, and 4 of Section 2.2.1 imply that the sequence of measures

\[
nP[a_n^{-1}(X, Y - b_n) \in (dx, dy)]
\]

converges vaguely [47, page 48] on \(([-\infty, \infty] \times [-M, \infty])\) as \( n \to \infty \) to the limit measure \( m_{2,\infty}(dx, dy) = e_0(dx) e^{-y} dy \) for some \( M > L \) (from Assumption 5 of Section 2.2.1) such that \(-M \) is a continuity point of \( [a_n^{-1}(Y - b_n)] \) for all \( n \).

**Remark 2.2.7.** Since the set of discontinuity points of the distribution functions of \( [a_n^{-1}(X - b_n)] \) for all \( n \) is countable, choice of such an \( M > L \) is not a problem.
Moreover, the $M$ in the two parts of the lemma (i) and (ii) may be chosen to be the same.

**Proof of Lemma 2.2.6.** We consider convergence of the measures evaluated on certain relatively compact regions which guarantee vague convergence.

**REGION 1:** $(x, \infty] \times [-\infty, y], x > -M, y \neq 0$. As $n \to \infty$,

\[
nP \left[ \frac{X - b_n}{a_n} > x, \frac{Y}{a_n} \leq y \right] = nP \left[ \frac{X - b_n}{a_n} > x \right] P \left[ \frac{Y}{a_n} \leq y \right] \frac{X - b_n}{a_n} > x \rightarrow e^{-x}1_{\{y > 0\}} = m_{1,\infty}((x, \infty] \times [-\infty, y])
\]

by Assumption 1 of Section 2.2.1 and Proposition 2.2.5.

**REGION 2:** $[-M, x] \times (y, \infty], x > -M, y \neq 0$. Since $-M$ is a continuity point of $\frac{X - b_n}{a_n}$ for all $n$, as $n \to \infty$,

\[
nP \left[ -M \leq \frac{X - b_n}{a_n} \leq x, \frac{Y}{a_n} > y \right] = nP \left[ -M < \frac{X - b_n}{a_n} \leq x, \frac{Y}{a_n} > y \right] \\
= nP \left[ \frac{X - b_n}{a_n} > -M, \frac{Y}{a_n} > y \right] - nP \left[ \frac{X - b_n}{a_n} > x, \frac{Y}{a_n} > y \right] \\
= nP \left[ \frac{X - b_n}{a_n} > -M \right] P \left[ \frac{Y}{a_n} > y \right] \frac{X - b_n}{a_n} > -M \\
- nP \left[ \frac{X - b_n}{a_n} > x \right] P \left[ \frac{Y}{a_n} > y \right] \frac{X - b_n}{a_n} > x \\
\rightarrow (e^{-M} - e^{-x})1_{\{y < 0\}} = m_{1,\infty}([-M, x] \times (y, \infty]),
\]

by Assumption 1 of Section 2.2.1 and Proposition 2.2.5.

Arguments for convergence on the following regions follow in a similar fashion using Proposition 2.2.5:

**REGION 3:** $(x, \infty] \times (y, \infty], x > -M, y \neq 0$,

**REGION 4:** $[-M, x] \times [-\infty, y], x > -M, y \neq 0$.

This concludes the proof of vague convergence on part (i).
The proof of part (ii) is similar but notice that if $c = 0$, we do not need Assumption 4. In this case, note that the limit measure $m_{2,\infty}(dx,dy)$ is a zero measure. Also note, using Assumptions 1 and 2, we get

$$nP \left[ \frac{Y - b_n}{a_n} \geq -M \right] \rightarrow ce^M = 0,$$

which is enough to prove the convergence in this case.

Now we prove Theorem 2.2.1.

Proof of Theorem 2.2.1. Choose $M$ as in Remark 2.2.7. We split $P(X + Y > b_n)$ as

$$P(X + Y > b_n) = P(X + Y > b_n, X > b_n - Ma_n) + P(X + Y > b_n, Y > b_n - Ma_n)$$

$$- P(X + Y > b_n, X > b_n - Ma_n, Y > b_n - Ma_n)$$

$$+ P(X + Y > b_n, X \leq b_n - Ma_n, Y \leq b_n - Ma_n). \quad (2.2.12)$$

Using Assumption 1 and (2.2.6), we get

$$nP(X + Y > b_n, X > b_n - Ma_n, Y > b_n - Ma_n) \leq nP(X > b_n - Ma_n, Y > b_n - Ma_n)$$

$$= nP(X > b_n - Ma_n) \frac{P(X > b_n - Ma_n, Y > b_n - Ma_n)}{P(X > b_n - Ma_n)} \rightarrow e^M = 0,$$

$$\quad (2.2.13)$$

since $b_n - Ma_n = b_n(1 - Ma_n/b_n) \rightarrow \infty$ [Remark 2.2.2 (1)]. Now, consider the convergence of the last term of (2.2.12) multiplied by $n$.

$$P(X + Y > b_n, X \leq b_n - Ma_n, Y \leq b_n - Ma_n) \leq nP(X > Ma_n, Y > Ma_n)$$

$$\sim \frac{P(X > Mf(b_n), Y > Mf(b_n))}{P(X > b_n)} \leq \frac{P(X > Lf(b_n), Y > Lf(b_n))}{P(X > b_n)} \rightarrow 0, \quad (2.2.14)$$

by Remark 2.2.2 (1) and Assumption 5 of Section 2.2.1.
To deal with the first term of (2.2.12) multiplied by \( n \), we first define a function \( T \) as \( T : [-M, \infty) \times [-\infty, \infty) \mapsto (\infty, \infty) \) by

\[
T(x, y) = \begin{cases} 
  x + y, & \text{if } y > -\infty, \\
  0, & \text{if } y = -\infty,
\end{cases}
\]

and hence

\[
nP(X + Y > b_n, X > b_n - Ma_n)
= nP(a_n^{-1}(X - b_n, Y) \in T^\rightarrow((0, \infty)) \cap \{(-M, \infty) \times [-\infty, \infty]\}). \tag{2.2.15}
\]

Note, that every set in the space \([-M, \infty) \times [-\infty, \infty]\) is relatively compact, and hence so is \( T^\rightarrow((0, \infty)) \cap \{(-M, \infty) \times [-\infty, \infty]\} = S \) (say). Also, since the limit measure \( m_{1,\infty}(\cdot) \) is concentrated on \([-M, \infty) \times \{0\}],

\[
m_{1,\infty}(\delta S) = m_{1,\infty}(\delta S \cap \{(-M, \infty) \times \{0\}\}) = m_{1,\infty}((0, 0)) = 0. \tag{2.2.16}
\]

Hence, using Lemma 2.2.6, (2.2.15) and (2.2.16), we get

\[
nP(X + Y > b_n, X > b_n - Ma_n) \to m_{1,\infty}(S) = 1. \tag{2.2.17}
\]

Similarly,

\[
nP(X + Y > b_n, Y > b_n - Ma_n) \to m_{2,\infty}(S) = c. \tag{2.2.18}
\]

Hence, using (2.2.12), (2.2.13), (2.2.14), (2.2.17) and (2.2.18), we get,

\[
\lim_{x \to \infty} \frac{P(X + Y > x)}{P(X > x)} = \lim_{n \to \infty} \frac{P(X + Y > b_n)}{P(X > b_n)} = \lim_{n \to \infty} nP(X + Y > b_n) = 1 + c,
\]

and we conclude our result. \( \Box \)

One immediate application of Theorem 2.2.1 is to the subexponential family of distributions denoted \( S \). The class \( MDA(\Lambda) \cap S \) has been studied in [23, page
and several sufficient conditions for belonging to this class are given in [26]. Corollary 2.2.8 gives an additional sufficient condition and follows directly from Theorem 2.2.1. Example 2.3.2 exhibits a distribution which satisfies the conditions of this Corollary.

**Corollary 2.2.8.** Suppose \( F_1(\cdot) \in MDA(\Lambda) \) with auxiliary function \( f(x) \) as described in Assumption 1 of Section 2.2.1. Suppose also \( \lim_{x \to \infty} f(x) = \infty \) and for some \( L > 0 \),

\[
\lim_{x \to \infty} \frac{\left[ \tilde{F}_1(L f(x)) \right]^2}{\tilde{F}_1(x)} = 0. \tag{2.2.19}
\]

Then, for \( X \) and \( Y \) i.i.d. with common distribution \( F_1(\cdot) \) we have, as \( x \to \infty \),

\[
P[X + Y > x] \sim 2P[X > x],
\]

and therefore, if \( F_1(\cdot) \) concentrates on \([0, \infty)\), \( F_1(\cdot) \in S \).

Following Remark 2.2.2(3), it is enough to check (2.2.19) with any \( \tilde{f}(x) \) satisfying \( \tilde{f}(x) \sim f(x) \). Note also it is natural to add the assumption \( f(x) \to \infty \), since if \( F_1(\cdot) \in MDA(\Lambda) \cap S \), then necessarily \( f(x) \to \infty \) [26].

**2.2.2 Asymptotic tail probability for the sum of more than two non-negative random variables**

Suppose among the risks \( X_1, X_2, \ldots, X_d \), there is no heavier tail than \( X_1 \) in the sense that

\[
\lim_{x \to \infty} \frac{\tilde{F}_i(x)}{\tilde{F}_1(x)} < \infty, \quad i = 2, \ldots, d.
\]

Assume \( X_1 \) satisfies Assumption 1 of Section 2.2.1 and that \( X_1, X_2, \ldots, X_d \) pairwise satisfy the Assumptions 3 and 4 of Section 2.2.1 with the auxiliary function \( f(\cdot) \)
of $X_1$. By this, we mean for all pairs $1 \leq i \neq j \leq d$, and for $t > 0$,
\[
\lim_{x \to \infty} \frac{P(X_j > tf(x), X_i > x)}{P(X_i > x)} = 0,
\]
which implies
\[
\lim_{x \to \infty} \frac{P(X_j > tf(x), X_i > x)}{P(X_i > x)} = 0. \tag{2.2.20}
\]

Also, suppose the risks $X_1, X_2, \ldots, X_d$ pairwise satisfy Assumption 5 of Section 2.2.1 with auxiliary function $f(\cdot)$ of $X_1$ so that for $1 \leq i < j \leq d$, there exists some $L_{ij} > 0$, such that either
\[
\lim_{x \to \infty} \frac{P(X_i > L_{ij}f(x), X_j > L_{ij}f(x))}{P(X_i > x)} = 0,
\]
or,
\[
\lim_{x \to \infty} \frac{P(X_i > L_{ij}f(x), X_j > L_{ij}f(x))}{P(X_j > x)} = 0.
\]
In either case, we have, for $1 \leq i < j \leq d$, for some $L_{ij} > 0$,
\[
\lim_{x \to \infty} \frac{P(X_i > L_{ij}f(x), X_j > L_{ij}f(x))}{P(X_1 > x)} = 0. \tag{2.2.21}
\]

Under the additional assumption of non-negativity, Theorem 2.2.1 can be extended to more than two risks.

**Corollary 2.2.9.** Assume $X_1, X_2, \ldots, X_d$ are non-negative random variables which pairwise satisfy Assumptions 3, 4, 5 of Section 2.2.1 with the auxiliary function $f(\cdot)$ of $X_1$. Moreover, the distribution of $X_1$ satisfies Assumption 1 of Section 2.2.1 and suppose
\[
\lim_{x \to \infty} \frac{P(X_i > x)}{P(X_1 > x)} = c_i \in [0, \infty), \quad i = 2, 3, \ldots, d. \tag{2.2.22}
\]

Define $S_j = X_1 + X_2 + \ldots + X_j$, $1 \leq j \leq d$ and we have, for $x \in \mathbb{R}$,
\[
\lim_{n \to \infty}nP(S_d > a_nx + b_n) = (1 + \sum_{i=2}^{d} c_i)e^{-x} \tag{2.2.23}
\]
and hence
\[
\lim_{x \to \infty} \frac{P(S_d > x)}{P(X_1 > x)} = (1 + \sum_{i=2}^{d} c_i). \tag{2.2.24}
\]
Remark 2.2.10.  1. **Asymptotic independence of the random variables:** Suppose for all \( i, c_i \in (0, \infty) \). Then for any \( 1 \leq i \neq j \leq d \), the pair \((X_i, X_j)\) is asymptotically independent by Remark 2.2.2(2). Since the random variables are pairwise asymptotically independent, they are also asymptotically independent [18, page 229].

2. **Non-negativity of random variables:** The only additional assumption added to the list in Section 2.2.1 is that the random variables are non-negative.

3. **Relaxation:** We have shown in (2.2.20) and (2.2.21) that pairwise satisfaction of Assumptions 3, 4, 5 of Section 2.2.1 implies that for \( 1 \leq i \neq j \leq d \), for \( t > 0 \),

\[
\lim_{x \to \infty} \frac{P(X_j > tf(x), X_i > x)}{P(X_i > x)} = 0,
\]

and for \( 1 \leq i < j \leq d \), there exists \( L_{ij} > 0 \),

\[
\lim_{x \to \infty} \frac{P(X_j > L_{ij}f(x), X_i > L_{ij}f(x))}{P(X_i > x)} = 0.
\]

We will show that actually these conditions are enough to get the desired conclusion.

**Proof.** We prove the result by induction under the relaxation Remark 2.2.10(3).

The base case of the induction for \( d = 2 \) is already proved in Theorem 2.2.1, so suppose the result is true for \( d = k \geq 2 \) and we have

\[
\lim_{n \to \infty} nP(S_k > a_n x + b_n) = (1 + \sum_{i=2}^{k} c_i) e^{-x} \quad (2.2.25)
\]

and

\[
\lim_{x \to \infty} \frac{P(S_k > x)}{P(X_1 > x)} = 1 + \sum_{i=2}^{k} c_i \quad (2.2.26)
\]
Therefore, we have
\[
\lim_{x \to \infty} \frac{P(X_{k+1} > x)}{P(S_k > x)} = \frac{c_{k+1}}{1 + \sum_{i=2}^k c_i} \in [0, \infty). \tag{2.2.27}
\]

We will use Theorem 2.2.1 with \( X = S_k \) and \( Y = X_{k+1} \). It remains to check the Assumptions in Theorem 2.2.1. For Assumption 1, note that \( S_k \) is tail equivalent to \( X_1 \) and use the fact that \( F_1(\cdot) \in MDA(\Lambda) \) is closed under tail equivalence. Assumption 2 is already checked in (2.2.27).

Note that from the induction hypothesis and asymptotic independence of \( X_i, i = 1, 2, \cdots, k \), \( P(S_k > x) \sim P(\bigcup_{i=1}^k (X_i > x)) \) and from the positivity of the risks \([S_k > x] \supseteq \bigcup_{i=1}^k [X_i > x] \). From these two facts, it easily follows that
\[
\lim_{x \to \infty} \frac{P((S_k > x) \cap (\bigcup_{i=1}^k (X_i > x))^c)}{P(S_k > x)} = 0. \tag{2.2.28}
\]

Since \( S_k \) and \( X_1 \) are tail equivalent, by [50], the auxiliary function \( \tilde{f}(\cdot) \) of \( S_k \) is asymptotically equal to the auxiliary function \( f(\cdot) \) of \( X_1 \), that is \( \lim_{x \to \infty} \tilde{f}(x)/f(x) = 1 \). Therefore, given \( \epsilon \in (0, 1) \), there exists \( T \) such that for all \( x > T, \tilde{f}(x) > \epsilon f(x) \). We now check Assumption 3. For \( t > 0, x > T \), using (2.2.28), as \( x \to \infty \),
\[
P([X_{k+1} > t\tilde{f}(x)|S_k > x] \leq P(X_{k+1} > t\epsilon f(x)|S_k > x)
\]
\[
= \frac{P(X_{k+1} > t\epsilon f(x), S_k > x)}{P(S_k > x)} \sim \frac{P(X_{k+1} > t\epsilon f(x), S_k > x, \bigcup_{i=1}^k [X_i > x])}{P(S_k > x)}
\]
\[
\leq \frac{P(X_{k+1} > t\epsilon f(x), \bigcup_{i=1}^k [X_i > x])}{P(S_k > x)} \leq \frac{\sum_{i=1}^k P(X_{k+1} > t\epsilon f(x), X_i > x)}{(1 + \sum_{i=2}^k c_i)P(X_1 > x)} \to 0
\]
by (2.2.20).

For Assumption 4, if \( c_{k+1} = 0 \), following Remark 2.2.2(4), there is no need to check Assumption 4. So, suppose \( c_{k+1} > 0 \). Then for any \( t > 0 \), as \( x \to \infty \),
\[
P([S_k > t\tilde{f}(x)|X_{k+1} > x] \leq P(S_k > t\epsilon f(x)|X_{k+1} > x)
\]

23
\[
\leq \sum_{i=1}^{k} P(X_i > t \epsilon f(x)/k | X_{k+1} > x)
= \sum_{i=1}^{k} \frac{P(X_i > t \epsilon f(x)/k, X_{k+1} > x)}{P(X_1 > x)} \frac{P(X_1 > x)}{P(X_{k+1} > x)} \to 0.
\]

For Assumption 5, we know from the assumptions in the statement of Corollary 2.2.9 that the random variables satisfy Assumption 5 of Section 2.2.1 pairwise with auxiliary function \( f(\cdot) \) of \( X_1 \). Thus, for \( 1 \leq i < j \leq d \), (2.2.21) holds. We check Assumption 5 with \( L = k L_{\text{max}} / \epsilon \), where \( L_{\text{max}} = \max_{1 \leq i \leq k} L_{i,k+1} \) (recall, equation (2.2.21)). Then, for sufficiently large \( x \), using \( \tilde{f}(\cdot) \) as the auxiliary function of \( S_k \),

\[
\frac{P(X_{k+1} > L_{i,k+1}f(x), S_k > L_{i,k+1}f(x))}{P(S_k > x)} \leq \frac{P(X_{k+1} > L_{i,k+1}f(x), \cup_{i=1}^k [X_i > L_{i,k+1}f(x)])}{P(S_k > x)} \leq \frac{P(X_{k+1} > L_{i,k+1}f(x), \cup_{i=1}^k [X_i > L_{i,k+1}f(x)])}{P(S_k > x)} \leq \frac{\sum_{i=1}^k P(X_{k+1} > L_{i,k+1}f(x), X_i > L_{i,k+1}f(x))}{P(S_k > x)} \leq \frac{\sum_{i=1}^k P(X_{k+1} > L_{i,k+1}f(x), X_i > L_{i,k+1}f(x))}{(1 + \sum_{i=2}^k c_i) P(X_1 > x)} \to 0
\]

by (2.2.21). This completes the induction proof. \( \square \)

### 2.3 Examples

This section shows a few of the many models that satisfy the Assumptions in Section 2.2.1. In all examples, both \( X \) and \( Y \) are non-negative random variables and it is straightforward to extend these examples to the \( d \)-dimensional case and show the assumptions of Corollary 2.2.9 are satisfied.

Our conditions are only sufficient and we exhibit one example where our conditions do not hold, but tail equivalence as in (2.2.3) holds true. Finding a
necessary and sufficient condition for the conclusion of Theorem 2.2.1 is still an open but subtle and difficult issue.

**Example 2.3.1.** Suppose \( X_1, X_2, X_3 \) are i.i.d. with common distribution \( F(\cdot) \), where for \( \alpha > 1 \),

\[
\tilde{F}(x) = \begin{cases} 
\exp\{-(\log x)^\alpha\}, & \text{if } x > 1, \\
1, & \text{if } x \leq 1.
\end{cases}
\]

Define

\[
X = X_1 \land X_2, \quad Y = X_2 \land X_3.
\]

It is easy to check \( X \) and \( Y \) are identically distributed with the common distribution \( \tilde{F}_1(\cdot) \), where

\[
\tilde{F}_1(x) = \exp(-2(\log x)^\alpha), \quad x > 1.
\]

It can be checked that \( F_1(\cdot) \) is a Von-Mises function; that is, it satisfies,

\[
\frac{\tilde{F}_1 F''_1}{(F'_1)^2} \rightarrow -1,
\]

a sufficient condition for \( F_1(\cdot) \in MDA(\Lambda) \), and

\[
f(x) = \frac{\tilde{F}_1(x)}{F'_1(x)} = \frac{x}{2\alpha(\log x)^{\alpha-1}}, \quad x > 1,
\]

serves as an auxiliary function [45, page 40]. Also, (2.2.1) is obvious and therefore, Assumption 1 of Section 2.2.1 is satisfied. Checking Assumption 2 is straightforward, so consider Assumption 3. Fix \( t > 0 \), recall \( f(x)/x \rightarrow 0 \), and note as \( x \rightarrow \infty \),

\[
\frac{P(X > x, Y > tf(x))}{P(X > x)} = \frac{P(X_1 > x, X_2 > x \lor tf(x), X_3 > tf(x))}{P(X_1 > x, X_2 > x)} \sim \frac{P(X_1 > x, X_2 > x, X_3 > tf(x))}{P(X_1 > x, X_2 > x)} = P(X_3 > tf(x)) \rightarrow 0,
\]

since \( f(x) \rightarrow \infty \). Assumption 4 is verified the same way. For Assumption 5, we have with \( L = 1 \),

\[
\frac{P(X > f(x), Y > f(x))}{P(X > x)} = \frac{P(X_1 > f(x), X_2 > f(x), X_3 > f(x))}{P(X_1 > x, X_2 > x)}
\]
\[
\frac{F(f(x))^3}{\bar{F}(x)^3} = \exp \left\{ - \left[ 3(\log f(x))^\alpha - 2\log x^\alpha \right] \right\} \\
= \exp \left\{ -2\log x^\alpha \left[ \frac{3}{2} \left( \frac{\log f(x)}{\log x} \right)^\alpha - 1 \right] \right\} \\
= \exp \left\{ -2\log x^\alpha \left[ \frac{3}{2} \left( 1 - \frac{\log(2\alpha(\log x)^{\alpha-1})}{\log x} \right)^\alpha - 1 \right] \right\} . 
\] 
\tag{2.3.1}

Since the exponent in (2.3.1) converges to \(-\infty\) as \(x \to \infty\), Assumption 5 is satisfied and this pair \((X,Y)\) satisfies the Assumptions in Section 2.2.1.

**Example 2.3.2.** Suppose \(X\) and \(Y\) are independent and identically distributed with common distribution \(F_1(\cdot)\), where for \(\alpha > 1\),

\[
\bar{F}_1(x) = \begin{cases} 
\exp(-(\log x)^\alpha) & \text{if } x > 1, \\
1 & \text{if } x \leq 1.
\end{cases}
\]

As in Example 2.3.1, one can check the subexponentiality condition (2.2.19) with 
\(L = 1\) and by Corollary 2.2.8, \(F_1(\cdot)\) is subexponential. Hence,

\[
P(X + Y > x) \sim 2P(X > x).
\]

**Example 2.3.3.** Suppose \(X \sim \text{Lognormal}(\mu, \sigma^2)\) and \(Y = e^{2\mu}/X\) so that \(X \overset{d}{=} Y\).

We check the Assumptions in Section 2.2.1 for the pair \((X,Y)\). The distribution \(\text{Lognormal}(\mu, \sigma^2)\) belongs to the maximal domain of attraction of the Gumbel distribution and its mean excess function \(e(x)\) has the form [23, page 147, 161]

\[
e(x) = \frac{\sigma^2 x}{\log x - \mu}(1 + o(1)).
\]

Also, (2.2.1) is obvious and so, Assumption 1 of Section 2.2.1 is true. Following Remark 2.2.2(3) and the form of \(e(x)\), we may choose the auxiliary function

\[
f(x) = \frac{\sigma^2 x}{\log x - \mu}.
\]

To verify Assumption 3, fix \(t > 0\), and note as \(x \to \infty\),

\[
\frac{P(X > x, Y > tf(x))}{P(X > x)} = \frac{P(X > x, e^{2\mu}/X > tf(x))}{P(X > x)} \to 0
\]
since \( f(x) \to \infty \). Assumption 4 is verified similarly. For Assumption 5, choose \( L = 1 \) and as \( x \to \infty \),

\[
\frac{P(X > f(x), Y > f(x))}{P(X > x)} = \frac{P(X > f(x), e^{2\mu / X} > f(x))}{P(X > x, Y > x)} \to 0.
\]

We conclude by Theorem 2.2.1,

\[
P(X + Y > x) \sim 2P(X > x).
\]

**Example 2.3.4.** Example 2.3.3 is a special case of a more general phenomenon. Suppose \( F_1(\cdot) \in MDA(\Lambda) \) with auxiliary function \( f(x) \) having the property

\[
\lim \inf_{x \to \infty} f(x) = \delta > 0. \tag{2.3.2}
\]

Assume that the support of \( F_1(\cdot) \) is a subset of \([0, \infty)\) and also satisfies the following conditions:

\[
x_0 = \sup \{x : F_1(x) < 1\} = \infty, \quad x_1 = \inf \{x : F_1(x) > 0\} = 0.
\]

Distributions satisfying these conditions include the exponential, gamma, log-normal. Define \( X = F_1^{-1}(U) \), and \( Y = F_1^{-1}(1 - U) \), where \( U \sim \text{Uniform}(0, 1) \). This pair \((X, Y)\) satisfies the Assumptions in Section 2.2.1.

Checking Assumption 2 is easy since \( X \) and \( Y \) are identically distributed. To verify Assumption 3, fix \( t > 0 \) and define \( \epsilon_t = F_1(t \delta / 2) \). Since \( x_1 = 0 \), we have \( \epsilon_t > 0 \). Then, for large \( x \) making \( f(x) > \delta / 2 \), we have

\[
\frac{P(X > x, Y > tf(x))}{P(X > x)} = \frac{P(U > F_1(x), 1 - U > F_1(tf(x)))}{P(X > x)} \leq \frac{P(U > F_1(x), 1 - U > \epsilon_t)}{P(X > x)} = \frac{P(U > F_1(x), U < 1 - \epsilon_t)}{P(X > x)} \to 0,
\]

since \( F_1(x) \to 1 \), and \( x_0 = \infty \). Assumption 4 is similarly verified. To verify Assumption 5, choose \( L \) such that \( F_1(L \delta / 2) > \frac{1}{2} \) and for \( x \) sufficiently large,

\[
\frac{P(X > Lf(x), Y > Lf(x))}{P(X > x)} \leq \frac{P(X > Lf(x), Y > \frac{L}{2})}{P(X > x)}
\]
\[ P(U > F_1(\frac{L_2}{2}), 1 - U > F_1(\frac{L_2}{2})) = \frac{P(X > x)}{P(X > x)} = 0. \]

Hence, \((X, Y)\) satisfy the Assumptions of Section 2.2.1 and by Theorem 2.2.1,

\[ P(X + Y > x) \sim 2P(X > x). \]

In this example, if \(\lim_{x \to \infty} f(x) = \infty\), we do not need the condition \(x_1 = 0\).

**Remark 2.3.5.** Note, in the previous two examples a comonotonic dependence structure is used. Also, note that in Example 2.3.4, the marginal distributions of \(X\) and \(Y\) are allowed not to be subexponential. This shows subexponentiality of the marginal distributions of \((X, Y)\) is not a required condition for the tail equivalence relation (2.2.3).

**Example 2.3.6.** Suppose \(X = \exp(X_1), Y = \exp(X_2)\), where \((X_1, X_2)\) is bivariate normal with correlation \(\rho \in [-1, 1]\). For simplicity, assume each \(X_i\) has mean \(\mu\) and variance \(\sigma^2 > 0\). This example is extensively considered in [5]. We have already considered the case \(\rho = -1\) in Example 2.3.3, so here we consider \(\rho \in (-1, 1)\).

Assumptions 1 and 2 of Section 2.2.1 are easily verified. Following the same reason as in Example 2.3.3, we take the auxiliary function to be

\[ f(x) = \frac{\sigma^2 x}{\log x - \mu}. \]

Observe,

\[
\frac{\log f(x) - \mu}{\sigma} = \frac{\log \left( \frac{\sigma^2 x}{\log x - \mu} \right) - \mu}{\sigma} = \frac{\log x - \mu}{\sigma} - \frac{1}{\sigma} \log \left( \frac{\log x - \mu}{\sigma^2} \right) = \left( \frac{\log x - \mu}{\sigma} \right) (1 + o(1)).
\]

For Assumption 3, we have for \(t > 0\), as \(x \to \infty\),

\[
\frac{P(X > x, Y > tf(x))}{P(X > x)} = \frac{P(X_1 > \log x, X_2 > \log tf(x))}{P(X_1 > \log x)}
\]
\[
\begin{align*}
\Phi \left( \frac{1}{\sqrt{2 \pi} (1+\rho)} \left( \log x + \log(tf(x)) - 2\mu \right) \right) & = \Phi \left( \frac{1}{\sqrt{2(1+\rho)}} \left( \log x - \mu \right) \right) \\
& = \Phi \left( \frac{2}{\sqrt{2(1+\rho)}} (1 + o(1)) \right) \\
& \rightarrow 0,
\end{align*}
\]

where we used (2.3.3) and the fact that $\Phi \in MDA(\Lambda)$ and therefore $\bar{\Phi}$ is $-\infty$-varying [45, page 53]. Note, $\rho < 1$ entails \( \frac{2}{\sqrt{2(1+\rho)}} > 1 \).

For Assumption 5, choose $L = 1$. As $x \to \infty$, we have using (2.3.3),
\[
\frac{P(X > f(x), Y > f(x))}{P(X > x)} = \frac{P(X_1 > \log f(x), X_2 > \log f(x))}{P(X_1 > \log x)} \\
\leq \frac{P(X_1 + X_2 > 2 \log f(x))}{P(X_1 > \log x)} = \frac{\Phi \left( \frac{2[\log f(x) - \mu]}{\sqrt{2 \pi^2 (1+\rho)}} \right)}{\bar{\Phi} \left( \frac{\log x - \mu}{\sigma} \right)} \\
= \frac{\Phi \left( \frac{2}{\sqrt{2(1+\rho)}} (1 + o(1)) \right)}{\bar{\Phi} \left( \frac{\log x - \mu}{\sigma} \right)} \to 0.
\]

**Example 2.3.7.** Let $X_1$ and $X_2$ be independent and identically distributed with the common distribution $D_1(\cdot) \in MDA(\Lambda)$, having auxiliary function $f_1(\cdot)$ satisfying (2.3.2) and infinite right end point. Also, suppose $D_2(\cdot) \in MDA(\Lambda)$ with auxiliary function $f_2(\cdot)$, concentrates on $[0, \infty)$ and satisfies the conditions in Example 2.3.4. Define $X$ and $Y$ as

\[
X = D_2^+(U) \land X_1, \quad Y = D_2^-(1-U) \land X_2,
\]

where $U$ is a uniformly distributed random variable on $(0, 1)$ which is independent of $(X_1, X_2)$.  

29
From Proposition 1.4 of [45, page 43], the distribution of $X$ belongs to the maximal domain of attraction of Gumbel with auxiliary function

$$f(x) = \frac{f_1(x)f_2(x)}{f_1(x) + f_2(x)}$$

Hence,

$$\limsup_{x \to \infty} \frac{1}{f(x)} \leq \limsup_{x \to \infty} \frac{1}{f_1(x)} + \limsup_{x \to \infty} \frac{1}{f_2(x)} < \infty,$$

and thus,

$$\liminf_{x \to \infty} f(x) > 0.$$ 

Also, note,

$$P(X > x) = P(U > D_2(x), X_1 > x) = P(U > D_2(x))P(X_1 > x) = \bar{D}_2(x)\bar{D}_1(x)$$

and

$$P(Y > x) = P(1-U > D_2(x), X_2 > x) = P(1-U > D_2(x))P(X_2 > x) = \bar{D}_2(x)\bar{D}_1(x).$$

Arguing as in Example 2.3.4, we can show that the pair $(X, Y)$ satisfy the assumptions in Section 2.2.1.

**Example 2.3.8.** Here is an example of a distribution for $(X, Y)$ where our assumptions are not satisfied, but the tail equivalence as in (2.2.3) is satisfied. Suppose $X$ and $Y$ are i.i.d. with common distribution $F_1(\cdot)$, where

$$\bar{F}_1(x) = \exp(-x^\alpha) \quad \alpha \in (0, 1), \quad x > 0.$$ 

This distribution is extensively studied in [54] and satisfies $F_1(\cdot) \in MDA(\Lambda) \cap S$. Since it is subexponential,

$$P(X + Y > x) \sim 2P(X > x).$$

However, this distribution does not satisfy Assumption 5 of Section 2.2.1.
Since $F_1(\cdot)$ is a Von-Mises function, we may take the auxiliary function to be

$$f(x) = \frac{\bar{F}_1(x)}{F_1(x)} = \frac{x^{1-\alpha}}{\alpha}.$$ 

Assumption 5 is not satisfied for any $L > 0$, since for any $L > 0$, as $x \to \infty$,

$$\frac{P(X > Lf(x), Y > Lf(x))}{P(X > x)} = \left[\frac{\bar{F}(Lf(x))}{\bar{F}(x)}\right]^2 = \exp\left(-2[Lf(x)]^{\alpha}\right) \exp(-x^{\alpha})$$

$$= \frac{\exp(-2\left(\frac{L}{\alpha}\right)^{\alphax^{(1-\alpha)}})}{\exp(-x^{\alpha})} = \exp\left(x^{\alpha} (1 - 2\left(\frac{L}{\alpha}\right)^{\alpha} x^{-\alpha^2})\right) \to \infty.$$ 

This also shows the criteria (2.2.19) for $F_1(\cdot) \in S$ is sufficient but not necessary.

2.4 Linear combinations of random variables with non-negative coefficients

This section studies linear combinations of risks $X, Y$ with non-negative coefficients. We consider two cases: (i) the distributions of $X$ and $Y$ are tail equivalent, and (ii) the distributions of $X$ and $Y$ lack tail equivalence. We explicitly give the asymptotic tail behavior of the linear combinations of risks in the tail equivalent case and also in one special case where tail equivalence is absent. We note that one cannot expect similar behavior in the two cases.

2.4.1 Tail equivalent cases

Linear combination of two random variables with non-negative coefficients

**Theorem 2.4.1.** Assume $(U, V)$ is a pair of random variables which satisfy Assumptions 1, 3, 4 and 5 of Section 2.2.1. Moreover, assume that Assumption 2
holds in the form
\[
\lim_{x \to \infty} \frac{P(V > x)}{P(U > x)} = c \in (0, \infty).
\] (2.4.1)

Define \( \hat{S}_2 = a_1 U + a_2 V \) and \( a_i \geq 0, i = 1, 2 \) and set \( m_2 = a_1 \vee a_2 \). Then, as \( x \to \infty \),
\[
P(\hat{S}_2 > x) \sim \frac{P(U > x)}{m_2} \left[ 1_{\{a_1 = m_2\}} + c 1_{\{a_2 = m_2\}} \right].
\]

We assume \( U \) and \( V \) are tail equivalent, that is the constant \( c \) cannot be 0 and hence both the marginal distributions belong to \( MDA(\Lambda) \), the maximal domain of attraction of the Gumbel. If \( \lim_{x \to \infty} P(V > x)/P(U > x) = 0 \), the asymptotic behavior of \( P(a_1 U + a_2 V > x) \) as \( x \to \infty \) can be different as illustrated in the following example.

**Example 2.4.2.** Assume \((U, V)\) are i.i.d. random variables with common distribution \( F_1(\cdot) \), which satisfy Assumptions 1, 3, 4 and 5 of Section 2.2.1. Define the two random vectors by \((U_1, V_1) = (U, \frac{1}{2} V)\) and \((U_2, V_2) = (U, \frac{1}{3} V)\). Both pairs \((U_1, V_1)\) and \((U_2, V_2)\) satisfy Assumptions 1, 3, 4 and 5 of Section 2.2.1. For both pairs \( c = 0 \), that is,
\[
\lim_{x \to \infty} \frac{P(V_1 > x)}{P(U_1 > x)} = 0 \quad \text{and} \quad \lim_{x \to \infty} \frac{P(V_2 > x)}{P(U_2 > x)} = 0.
\]

Since, \((U, V)\) satisfies the Assumptions of Theorem 2.4.1, we have as \( x \to \infty \),
\[
P(3U_1 + 10V_1 > x) = P(3U + 2V > x) \sim P(3U > x) = P(3U_1 > x),
\]
and
\[
P(3U_2 + 10V_2 > x) = P(3U + 5V > x) \sim P(5V > x) = P(10V_2 > x).
\]
This example illustrates we cannot expect Theorem 2.4.1 to hold for the case \( c = 0 \).
We now turn to the proof of Theorem 2.4.1.

Proof. The case \(a_1 = a_2\) is resolved by Theorem 2.2.1 since

\[
P(a_1(U + V) > x) = P(U + V > \frac{x}{a_1}) \sim (1 + c)P(U > \frac{x}{a_1}).
\]

So the interesting cases are \(a_1 > a_2\) and \(a_1 < a_2\) and for the following, assume \(a_1 > a_2\), the other case being similar.

There is nothing to prove if \(a_2 = 0\), so assume \(a_1 > a_2 > 0\) which makes \(m_2 = a_1\). It suffices to check the Assumptions in Section 2.2.1 for \(X = U\) and \(Y = a_2V/a_1\). For this definition of \(X, Y\), we have

\[
\lim_{x \to \infty} \frac{P(Y > x)}{P(X > x)} = \lim_{x \to \infty} \frac{P(a_2V/a_1 > x)}{P(U > x)} = \lim_{x \to \infty} \frac{P(V > a_1x/a_2)}{P(U > x)} = 0.
\]  

(2.4.2)

The last equality is true from (2.4.1) and the fact that the tail of any distribution in \(MDA(\Lambda)\) is \(-\infty\)-varying [45, page 53]. From Theorem 2.2.1 and (2.4.2), we get, as \(x \to \infty\),

\[
P(a_1U + a_2V > x) = P(a_1(U + a_2V/a_1) > x) = P(U + a_2V/a_1 > x/a_1) = P(X + Y > x/a_1) \sim P(U > x/a_1) = P(U > \frac{x}{m_2})[1_{a_1 = m_2} + c1_{a_2 = m_2}].
\]

Assumption 1 of Section 2.2.1 is assumed in the statement of Theorem 2.4.1 and Assumption 2 was verified in (2.4.2). For Assumption 3, note that \(U \in MDA(\Lambda)\) and suppose \(f(\cdot)\) is the auxiliary function of the distribution of \(U\). By hypothesis, for \(t > 0\),

\[
\lim_{x \to \infty} P(|V| > tf(x)|U > x) = 0,
\]

(2.4.3)

and therefore, using (2.4.3),

\[
\lim_{x \to \infty} P(a_2|V|/a_1 > tf(x)|U > x) = \lim_{x \to \infty} P(|V| > a_1tf(x)/a_2|U > x) = 0.
\]
Remark 2.2.2(4) implies we do not need to verify Assumption 4, so we check Assumption 5. For this we have, as $x \to \infty$,

\[
\frac{P(a_2V/a_1 > Lf(x), U > Lf(x))}{P(U > x)} = \frac{P(V > a_1Lf(x)/a_2, U > Lf(x))}{P(U > x)} \leq \frac{P(V > Lf(x), U > Lf(x))}{P(U > x)} \to 0.
\]

This proves the case $a_1 > a_2$. □

**Linear Combination of more than two random variables with non-negative coefficients**

**Corollary 2.4.3.** Assume $X_1, X_2, \ldots, X_d$ are non-negative random variables which pairwise satisfy Assumptions 3, 4, 5 of Section 2.2.1. Further suppose the distribution of $X_1$ satisfies Assumption 1 of Section 2.2.1 and that

\[
\lim_{x \to \infty} \frac{P(X_i > x)}{P(X_1 > x)} = c_i \in (0, \infty), \quad i = 1, 2, \ldots, d. \tag{2.4.4}
\]

Set $c_1 = 1$ and define for $d > 1$, $\hat{S}_d = a_1X_1 + a_2X_2 + \ldots + a_dX_d$, for $a_i \geq 0$, $i = 1, 2, \ldots, d$. Also, define

\[
m_d = \sqrt[d]{a_i} \quad \text{and} \quad N_d = \sum_{\{1 \leq i \leq d : a_i = m_d\}} c_i.
\]

Then

\[
P(\hat{S}_d > x) \sim N_dP(X_1 > \frac{x}{m_d}), \quad x \to \infty.
\]

This result is consistent with the case where $X_1, X_2, \ldots, X_d$ are i.i.d. with common distribution in $MDA(\Lambda) \cap S$; see [13].

The random variables $X_1, X_2, \ldots, X_d$ are tail equivalent and satisfy Assumption 3 of Section 2.2.1 pairwise. Therefore Remark 2.2.2(2) implies pairwise
asymptotic independence and hence, by [45, page 291], \(X_1, \ldots, X_d\) are asymptotically independent.

In the special case that the random variables are identically distributed, we get \(N_d = |\{1 \leq i \leq d : a_i = m_d\}|\), where \(|\cdot|\) is the size of a set.

**Remark 2.4.4.** It is possible to prove Corollary 2.4.3 using Corollary 2.2.9. However, in the proof it is usually difficult to verify Assumption 4 of Section 2.2.1. Note, a similar problem is avoided carefully in the proof of Theorem 2.4.1 through the help of Remark 2.2.2(4). Though a remark similar to Remark 2.2.2(4) could also be made for Corollary 2.2.9, it is notationally inconvenient. So, to avoid this notational difficulty, Theorem 2.4.1 is used for the proof.

**Proof.** Proceeding by induction, note the base case for \(d = 2\) is proved in Theorem 2.4.1. As an induction hypothesis, suppose the result is true for \(d = k\), so, as \(x \to \infty\),

\[
P(\hat{S}_k > x) \sim N_k P(X_1 > \frac{x}{m_k}) \sim \frac{N_k}{c_{k+1}} P(X_{k+1} > \frac{x}{m_k}).
\]

To prove the result for \(d = k + 1\), notice,

\[
m_{k+1} = m_k \lor a_{k+1},
\]

and

\[
N_{k+1} = [c_{k+1} I_{\{a_{k+1} = m_{k+1}\}} + N_k I_{\{m_k = m_{k+1}\}}],
\]

so that

\[
\frac{N_{k+1}}{c_{k+1}} = \left[ I_{\{a_{k+1} = m_{k+1}\}} + \frac{N_k}{c_{k+1}} I_{\{m_k = m_{k+1}\}} \right].
\]

By the induction hypothesis,

\[
\lim_{x \to \infty} \frac{P(m_k^{-1} \hat{S}_k > x)}{P(X_{k+1} > x)} = \lim_{x \to \infty} \frac{P(m_k^{-1} \hat{S}_k > x)}{P(X_1 > x)} \frac{P(X_1 > x)}{P(X_{k+1} > x)} = \frac{N_k}{c_{k+1}}.
\]
If we prove the assumptions in Theorem 2.4.1 are valid with $U = X_{k+1}$ and $V = m_k^{-1} \hat{S}_k$, then, Theorem 2.4.1, (2.4.5), (2.4.6) and (2.4.7) imply, as $x \to \infty$,

$$P(\hat{S}_{k+1} > x) = P(a_{k+1}X_{k+1} + m_k \hat{S}_k > x) \sim \frac{N_{k+1}}{c_{k+1}} P(X_{k+1} > x) \sim N_{k+1} P(X_1 > x),$$

and by induction, our result holds for all $d \geq 2$.

Assumption 1 of Section 2.2.1 is assumed for $X = X_{k+1}$. For (2.4.1), consider that on the one hand,

$$N_k = \sum_{\{1 \leq i \leq k : a_i = m_k\}} c_i \geq \bigwedge_{i=1}^k c_i > 0 \quad (2.4.8)$$

and on the other,

$$N_k = \sum_{\{1 \leq i \leq k : a_i = m_k\}} c_i \leq k \bigvee_{i=1}^k c_i < \infty, \quad (2.4.9)$$

and therefore the limit in (2.4.7) satisfies $N_k/c_{k+1} \in (0, \infty)$.

To verify the rest of the assumptions of Section 2.2.1, we first note an important fact. Suppose two random variables $U$ and $V$ are tail equivalent and both belong to $MDA(\Lambda)$. If $f(\cdot)$, $\tilde{f}(\cdot)$ are the auxiliary functions of $U$ and $V$ respectively, then $f(x) \sim \tilde{f}(x)$, as $x \to \infty$; see [49, 50]. Since, in the present case, all the random variables are tail equivalent, Remark 2.2.2(3) implies we can work with the auxiliary function of any one of them, say $X_{k+1}$. So, $X_1, X_2, \ldots, X_d$ satisfy Assumptions 3, 4 and 5 of Section 2.2.1 pairwise with the auxiliary function $f(\cdot)$ of $X_{k+1}$. That is, for $1 \leq i \neq j \leq d$, and any $t > 0$,

$$\lim_{x \to \infty} P(X_j > tf(x)|X_i > x) = 0 \quad (2.4.10)$$

and for $1 \leq i < j \leq d$, for some $L_{ij} > 0$

$$\frac{P(X_i > L_{ij} f(x), X_j > L_{ij} f(x))}{P(X_i > x)} = 0. \quad (2.4.11)$$
To verify Assumption 3, observe for $t > 0$, that as $x \to \infty$,

\[
P(|m_k^{-1} \hat{S}_k| > tf(x)|X_{k+1} > x)
\leq P(a_1X_1 + a_2X_2 + \cdots + a_kX_k > m_k tf(x)|X_{k+1} > x)
\leq P(\bigcup_{i=1}^k [a_iX_i > m_k tf(x)/k]|X_{k+1} > x)
\leq \sum_{i=1}^k P(X_i > a_i^{-1} m_k tf(x)/k|X_{k+1} > x) \leq \sum_{i=1}^k P(X_i > tf(x)/k|X_{k+1} > x) \to 0,
\]

by (2.4.10). For Assumption 4, note,

\[
\lim_{x \to \infty} \frac{P(m_k^{-1} \hat{S}_k > x)}{P(X_1 > x)} = N_k, \tag{2.4.12}
\]

and for $1 \leq i \leq k$,

\[
\lim_{x \to \infty} \frac{P((m_k^{-1} \hat{S}_k > x) \cap (m_k^{-1} a_iX_i > x))}{P(X_1 > x)} = \lim_{x \to \infty} \frac{P(m_k^{-1} a_iX_i > x)}{P(X_1 > x)} = c_i 1_{\{a_i=m_i\}}. \tag{2.4.13}
\]

The first equality uses the assumption that $X_i$’s are non-negative. The second equality is true from (2.4.4) and the fact that the tail of any distribution in the maximal domain of attraction of Gumbel (MDA($\Lambda$)) is $-\infty$-varying [45, page 53].

Now, for $1 \leq i < j \leq k$, using (2.2.6),

\[
\lim_{x \to \infty} \frac{P((m_k^{-1} \hat{S}_k > x) \cap (m_k^{-1} a_iX_i > x) \cap (m_k^{-1} a_jX_j > x))}{P(X_1 > x)}
\leq \lim_{x \to \infty} \frac{P((m_k^{-1} a_iX_i > x) \cap (m_k^{-1} a_jX_j > x))}{P(X_1 > x)}
\leq \lim_{x \to \infty} \frac{P((X_i > x) \cap (X_j > x))}{P(X_1 > x)} = 0.
\]

Therefore, using (2.4.13),

\[
\lim_{x \to \infty} \frac{P((m_k^{-1} \hat{S}_k > x) \cap (\bigcup_{i=1}^k (m_k^{-1} a_iX_i > x)))}{P(X_1 > x)}
= \lim_{x \to \infty} \sum_{i=1}^k \frac{P((m_k^{-1} \hat{S}_k > x) \cap (m_k^{-1} a_iX_i > x))}{P(X_1 > x)} = N_k. \tag{2.4.14}
\]

37
From (2.4.12) and (2.4.14) it follows that

\[ \lim_{x \to \infty} \frac{P((m_k^{-1} \hat{S}_k > x) \cap (\cup_{i=1}^k (m_k^{-1}a_iX_i > x))^c)}{P(X_1 > x)} = 0, \]

and this, along with (2.4.4) and (2.4.7) gives

\[ \lim_{x \to \infty} \frac{P((m_k^{-1} \hat{S}_k > x) \cap (\cup_{i=1}^k (m_k^{-1}a_iX_i > x))^c)}{P(\hat{S}_k > x)} = 0. \tag{2.4.15} \]

Now, we check Assumption 4. For \( t > 0 \), as \( x \to \infty \),

\[ P(|X_{k+1}| > tf(x)|m_k^{-1} \hat{S}_k > x) = \frac{P(X_{k+1} > tf(x), m_k^{-1} \hat{S}_k > x)}{P(m_k^{-1} \hat{S}_k > x)} \]

\[ \sim \frac{P(X_{k+1} > tf(x), m_k^{-1} \hat{S}_k > x, \cup_{i=1}^k (m_k^{-1}a_iX_i > x))}{P(m_k^{-1} \hat{S}_k > x)} \]

\[ \leq \frac{P(X_{k+1} > tf(x), \cup_{i=1}^k (m_k^{-1}a_iX_i > x))}{P(\hat{S}_k > m_kx)} \]

\[ \leq \frac{\sum_{i=1}^k P(X_{k+1} > tf(x), m_k^{-1}a_iX_i > x)}{P(\hat{S}_k > m_kx)}, \]

where we have used (2.4.15). Using our induction hypothesis, we get that the quantity above is asymptotically equivalent to

\[ \sim \frac{\sum_{i=1}^k P(X_{k+1} > tf(x), m_k^{-1}a_iX_i > x)}{N_k P(X_1 > x)} \]

\[ \leq \frac{\sum_{i=1}^k P(X_{k+1} > tf(x), X_i > x)}{N_k P(X_1 > x)} \]

\[ = \frac{\sum_{i=1}^k P(X_{k+1} > tf(x), X_i > x)}{P(X_i > x)} \frac{P(X_i > x)}{N_k P(X_1 > x)} \to 0, \]

by (2.4.10).

For Assumption 5, let, \( L = kL_{\max} \) where \( L_{\max} = \max_{1 \leq i \leq k} L_{i,k+1} \) (recall, equation (2.4.11)). Then using (2.4.7), (2.4.11) and (2.4.15), we have

\[ \frac{P(X_{k+1} > kL_{\max}f(x), m_k^{-1} \hat{S}_k > kL_{\max}f(x))}{P(X_{k+1} > x)} \]

\[ \sim \frac{P(X_{k+1} > kL_{\max}f(x), m_k^{-1} \hat{S}_k > x, \cup_{i=1}^k (m_k^{-1}a_iX_i > x))}{P(X_{k+1} > x)} \]

38
\[
\begin{align*}
&\leq \frac{P(X_{k+1} > kL_{\text{max}}f(x), \bigcup_{i=1}^{k} \{m_i^{-1}a_iX_i > L_{\text{max}}f(x)\})}{P(X_{k+1} > x)} \\
&\leq \sum_{i=1}^{k} \frac{P(X_{k+1} > L_{i,k+1}f(x), m_i^{-1}a_iX_i > L_{i,k+1}f(x))}{P(X_{k+1} > x)} \\
&\leq \sum_{i=1}^{k} \frac{P(X_{k+1} > L_{i,k+1}f(x), X_i > L_{i,k+1}f(x))}{P(X_{k+1} > x)} \rightarrow 0.
\end{align*}
\]

\[\square\]

### 2.4.2 One special case where the distributions are possibly not tail equivalent

**Theorem 2.4.5.** Assume \(Y_1, Y_2, \ldots, Y_d\) are identically distributed non-negative random variables. Also, assume \(a_i, \beta_i \geq 0, i = 1, 2, \ldots, d\). For \(d \geq 1\), define \(\hat{S}_d = a_1 Y_1^{\beta_1} + a_2 Y_2^{\beta_2} + \ldots + a_d Y_d^{\beta_d}\) and set

\[
\beta = \bigvee_{i=1}^{d} \beta_i, \quad q_d = \bigvee_{\{1 \leq i \leq d : \beta_i = \beta\}} a_i, \\
J_d = |\{1 \leq i \leq d : \beta_i = \beta, a_i = q_d\}|
\]

where \(|\cdot|\) denotes the size of the set. Suppose \(q_d Y_1^{\beta}, q_d Y_2^{\beta}, \ldots, q_d Y_d^{\beta}\) pairwise satisfy Assumptions 3, 4 and 5 of Section 2.2.1 and that the distribution of \(q_d Y_1^{\beta}\) satisfies Assumption 1 of Section 2.2.1 where its auxiliary function \(f(x)\) satisfies the additional condition that \(f(x) \to \infty\), as \(x \to \infty\). Then,

\[P(\hat{S}_d > x) \sim J_d P(Y_1^{\beta} > \frac{x}{q_d}).\]

**Remark 2.4.6.** If \(\beta_1 > \beta_2\), then \(Y_1^{\beta_1}\) and \(Y_2^{\beta_2}\) are not tail equivalent. Note, in this case, the asymptotic approximation of \(P(a_1 Y_1^{\beta_1} + a_2 Y_2^{\beta_2} > x)\) does not depend on \(a_2\), which is different from the asymptotic result observed in Theorem 2.4.1.
Theorem 2.4.5 shows different tail behavior from the tail equivalent cases but follows the paradigm that only the heaviest tails matter. It also shows that Theorem 1 of [5] is a special case of a more general phenomenon. Let 
\((Z_1, Z_2, \ldots, Z_d) \sim N(0, \Sigma)\), where

\[ \Sigma = (\rho_{ij}), \quad \rho_{ii} = 1, \quad \forall i, \quad \rho_{ij} < 1 \quad \text{for} \quad 1 < i < j \leq d. \]

Let, 
\((Y_1, Y_2, \ldots, Y_d) \sim (\exp(Z_1), \exp(Z_2), \ldots, \exp(Z_d))\). Clearly,

\[ a_i Y_i^{\beta_i} \sim \text{Lognormal}(\log a_i, \beta_i^2). \]

From Example 2.3.6, 
\((q_d Y_1^{q_d}, q_d Y_2^{q_d}, \ldots, q_d Y_d^{q_d})\) satisfy the assumptions of Theorem 2.4.5, where \(q_d, \beta\) have the same meaning as in Theorem 2.4.5. Also, 
\((W_1, W_2, \ldots, W_d) = (a_1 Y_1^{a_1}, a_2 Y_2^{a_2}, \ldots, a_d Y_d^{a_d})\) satisfies the assumptions of Theorem 1 of [5]. The results of that theorem and Theorem 2.4.5 match.

**Proof.** Without loss of generality, assume \(\beta_1 = \beta\) and \(a_1 = q_d\). Also, assume \(a_i > 0\) for \(i = 1, 2, \ldots, d\). Denote,

\[ X_i = a_i Y_i^{\beta_i} \quad i = 1, 2, \ldots, d. \]

To start, suppose for some \(i \in \{2, \ldots, d\}, \beta_i < \beta\). Then, for large \(x\), \([a_i Y_i^{\beta_i} > x] \subseteq \left[ \frac{q_d}{2} Y_i^{\beta_i} > x \right]\), and hence for large \(x\),

\[ P(a_i Y_i^{\beta_i} > x) \leq P(\frac{q_d}{2} Y_i^{\beta_i} > x) = P(q_d Y_1^{\beta} > 2x). \]

Then,

\[ c_i = \lim_{x \to \infty} \frac{P(X_i > x)}{P(X_1 > x)} = \lim_{x \to \infty} \frac{P(a_i Y_i^{\beta_i} > x)}{P(q_d Y_i^{\beta} > x)} \leq \lim_{x \to \infty} \frac{P(q_d Y_1^{\beta} > 2x)}{P(q_d Y_1^{\beta} > x)} = 0. \quad (2.4.16) \]

Next, suppose for some \(i \in \{2, \ldots, d\}, \beta_i = \beta, a_i < q_d\). Then,

\[ c_i = \lim_{x \to \infty} \frac{P(X_i > x)}{P(X_1 > x)} = \lim_{x \to \infty} \frac{P(a_i Y_i^{\beta} > x)}{P(q_d Y_1^{\beta} > x)} = \lim_{x \to \infty} \frac{P(q_d Y_1^{\beta} > \frac{q_d X}{a_i})}{P(q_d Y_1^{\beta} > x)} = 0. \quad (2.4.17) \]
In both the equations (2.4.16) and (2.4.17), the last equalities are true from the fact that the tail of any distribution in the maximal domain of attraction of the Gumbel is $-\infty$-varying [45, page 53].

Finally, suppose for some $i \in \{2, \ldots, d\}, \beta_i = \beta, a_i = q_d$.

$$c_i = \lim_{x \to \infty} \frac{P(X_i > x)}{P(X_1 > x)} = \lim_{x \to \infty} \frac{P(Y_{1}^{\beta} > \frac{x}{q_d})}{P(Y_{1}^{\beta} > \frac{x}{q_d})} = \lim_{x \to \infty} \frac{P(Y_{1}^{\beta} > \frac{x}{q_d})}{P(Y_{1}^{\beta} > \frac{x}{q_d})} = 1. \quad (2.4.18)$$

It suffices to check the assumptions in Corollary 2.2.9 with this set of $X_1, X_2, \ldots, X_d$, since then Corollary 2.2.9 and (2.4.16), (2.4.17), (2.4.18) would imply, as $x \to \infty$,

$$P(\hat{S}_d > x) \sim (1 + \sum_{i=2}^{d} c_i)P(X_1 > x) \sim J_dP(X_1 > x) = J_dP(Y_{1}^{\beta} > \frac{x}{q_d}).$$

Assumption 1 of Section 2.2.1 is assumed for $X = q_dY_{1}^{\beta}$ in the statement of Theorem 2.4.5 and (2.2.22) is already shown in (2.4.16), (2.4.17) and (2.4.18). For Assumptions 3 and 4 of Section 2.2.1, proceed as follows. By hypothesis, we know that $X_1$ belongs to the maximal domain of attraction of the Gumbel distribution. Let $f(\cdot)$ be the auxiliary function corresponding to the distribution of $X_1$. By hypothesis, we know, for $t > 0$, for $1 \leq i \neq j \leq d$,

$$\lim_{x \to \infty} P(q_dY_{j}^{\beta} > tf(x)|q_dY_{i}^{\beta} > x) = 0. \quad (2.4.19)$$

Using Remark 2.2.10(3), it is enough to show

$$\lim_{x \to \infty} \frac{P(X_j > tf(x), X_i > x)}{P(X_1 > x)} = 0,$$

and to see this, note that since $f(x) \to \infty$, for large $x$ and for all $t > 0$, $[X_j > tf(x), X_i > x] \subseteq [q_dY_{j}^{\beta} > tf(x), q_dY_{i}^{\beta} > x]$. Hence, by (2.4.19)

$$\lim_{x \to \infty} \frac{P(X_j > tf(x), X_i > x)}{P(X_1 > x)} \leq \lim_{x \to \infty} \frac{P(q_dY_{j}^{\beta} > tf(x), q_dY_{i}^{\beta} > x)}{P(q_dY_{i}^{\beta} > x)} = 0.$$
For Assumption 5, using Remark 2.2.10(3), we show, for some $L > 0$,

$$\lim_{x \to \infty} \frac{P(X_j > Lf(x), X_i > Lf(x))}{P(X_1 > x)} = 0.$$ 

By hypothesis, we know, for all $1 \leq i < j \leq d$, there exists some $L_{ij} > 0$,

$$\lim_{x \to \infty} \frac{P(q_d Y_j^\beta > L_{ij} f(x), q_d Y_i^\beta > L_{ij} f(x))}{P(q_d Y^\beta > x)} = 0. \quad (2.4.20)$$

Also, note that $[X_j > L_{ij} f(x), X_i > L_{ij} f(x)] \subseteq [q_d Y_j^\beta > L_{ij} f(x), q_d Y_i^\beta > L_{ij} f(x)]$ for large $x$, since $f(x) \to \infty$. Hence, by (2.4.20),

$$\lim_{x \to \infty} \frac{P(X_j > L_{ij} f(x), X_i > L_{ij} f(x))}{P(X_1 > x)} \leq \lim_{x \to \infty} \frac{P(q_d Y_j^\beta > L_{ij} f(x), q_d Y_i^\beta > L_{ij} f(x))}{P(q_d Y^\beta_1 > x)} = 0.$$

\[\square\]

### 2.5 An optimization problem

Suppose we have a portfolio consisting of $d$ financial instruments. The risk per unit of the $i$-th instrument is $X_i$. The goal is to earn revenue $\$L$. Assume each unit of the $i$-th instrument earns $\$l_i$ over the chosen time horizon. Subject to earnings being at least $\$L$, how many units of each instrument, $a_1, a_2, \ldots, a_d$, should be used to build the portfolio, so that the probability that the total portfolio risk $a_1 X_1 + a_2 X_2 + \ldots + a_d X_d$ exceeds some fixed large threshold $x$, is minimal?

Thus, consider the following optimization problem:

$$\min_{\{a_1, \ldots, a_d\}} P \left[ \sum_{i=1}^d a_i X_i > x \right]$$

s.t. $a_1 l_1 + a_2 l_2 + \cdots + a_d l_d \geq L$

$a_i \geq 0$, $i = 1, 2, \ldots, d$.  

42
For a more general case, consider the following optimization problem:

\[
\min_{[a_1, \ldots, a_d]} P \left[ \sum_{i=1}^{d} a_i X_i > x \right]
\]

s.t. \( h(a_1, a_2, \ldots, a_d) \geq L, \)

\( a_i \geq 0, \ i = 1, 2, \ldots, d. \)

### 2.5.1 An approximate solution

Suppose \( X_1, X_2, \ldots, X_d \) satisfy the assumptions of Corollary 2.4.3. Even with these assumptions, exact solution of the optimization problem is difficult. An obvious way to obtain an approximate solution to the optimization problem is to assume that the threshold \( x \) is large and use the asymptotic approximation of \( P(a_1 X_1 + a_2 X_2 + \ldots + a_d X_d > x) \) from Corollary 2.4.3, hoping that the solution of the resulting optimization problem is close to the actual optimal value. So, using the notation of Corollary 2.4.3, we solve the following optimization problem:

\[
\min_{[\hat{a}_1, \ldots, \hat{a}_d]} N_d P(X_1 > \frac{x}{\hat{m}_d})
\]

s.t. \( h(\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_d) \geq L, \)

\( \hat{a}_i \geq 0, \ i = 1, 2, \ldots, d. \)

Suppose \( \hat{a}_1, \hat{a}_2, \ldots, \hat{a}_d \) and \( \bar{a}_1, \bar{a}_2, \ldots, \bar{a}_d \) are two feasible solutions for the given set of constraints. Set

\[
\hat{m}_d = \bigvee_{i=1}^{d} \hat{a}_i, \quad \hat{N}_d = \sum_{\{1 \leq i \leq d : \hat{a}_i = \hat{m}_d\}} c_i
\]

\[
\bar{m}_d = \bigvee_{i=1}^{d} \bar{a}_i, \quad \bar{N}_d = \sum_{\{1 \leq i \leq d : \bar{a}_i = \bar{m}_d\}} c_i
\]
If \( \hat{m}_d > \bar{m}_d \), then since, \( P[X_1 \leq x] \in MDA(\Lambda) \), as \( x \to \infty \),
\[
\frac{P(X_1 > x/\hat{m}_d)}{P(X_1 > x/\bar{m}_d)} \to \infty.
\]

Now, since both \( \hat{N}_d, \bar{N}_d \in \left[ \bigwedge_{i=1}^d c_i, d \bigvee_{i=1}^d c_i \right], \) we have as \( x \to \infty \),
\[
\frac{\hat{N}_d P(X_1 > x/\hat{m}_d)}{\bar{N}_d P(X_1 > x/\bar{m}_d)} \to \infty.
\]

So, we hope that \( \bar{a}_1, \bar{a}_2, \ldots, \bar{a}_d \) is a better feasible solution for the optimization problem.

Thus, values of \( a_1, a_2, \ldots, a_d \) which approximately solve the above optimization problem can be computed by solving the following two optimization problems in sequence.

(i) First solve
\[
\min_{\{a_1, \ldots, a_d\}} m_d = \max\{a_1, a_2, \ldots, a_d\}
\]
\[
\text{s.t. } h(a_1, a_2, \ldots, a_d) \geq L,
\]
\[
a_i \geq 0, \quad i = 1, 2, \ldots, d.
\]

(ii) Suppose the best choice of \( a_1, a_2, \ldots, a_d \) gives \( m \) as the value of the objective function for the optimization problem in (i). Then we solve
\[
\min_{\{a_1, \ldots, a_d\}} N_d = \sum_{\{1 \leq i \leq d; a_i = m\}} c_i
\]
\[
\text{s.t. } h(a_1, a_2, \ldots, a_d) \geq L,
\]
\[
\max\{a_1, a_2, \ldots, a_d\} = m,
\]
\[
a_i \geq 0 \quad i = 1, 2, \ldots, d.
\]
2.5.2 A special case

The motivating case is that $h(\cdot)$ is a linear function with positive coefficients of the form

$$h(a_1, a_2, \ldots, a_d) = a_1 l_1 + a_2 l_2 + \ldots + a_d l_d.$$ 

The approximate solution using the asymptotic form of $P[\sum_{i=1}^d a_i X_i > x]$ is

$$a_1 = a_2 = \ldots = a_d = L / (l_1 + l_2 + \ldots + l_d).$$

This leads to $m = L / (l_1 + l_2 + \ldots + l_d)$ and $N_d = \sum_{i=1}^d c_i$.

2.6 Simulation studies

We carried out some simulation studies to check for fixed large thresholds the accuracy of the asymptotic approximation in Theorem 2.2.1 and also to check how good is the approximate solution for the optimization problem. As expected, in some cases the approximation works well whereas in others it performs poorly which suggests caution about using the asymptotic results for numerical purposes. Simulation also suggests that the approximate solution of the optimization problem works well in cases where simulation studies suggest that the approximation is good for fixed large thresholds. One particular model studied, Example 2.3.6 with $\mu = 0, \sigma = 1$, is noted here to illustrate the point. We varied $\rho$ and observed the asymptotic behavior of the sum of the risks.
2.6.1 Where is the approximation good?

To test the approximation for $P(X + Y > x)$, we have to find good simulation estimates of the probabilities $P(X + Y > x)$. This, however is not easy, especially in the case when the marginal distributions of the risks $X$ and $Y$ are subexponential and is still a topic of current research in the simulation community. The approach usually taken in these cases is Conditional Monte Carlo [4, page 173]. So, this method is used to compute $P(X + Y > x)$ and the simulation estimates are compared with the theoretical approximations.

The simulation of $P(X + Y > x)$ uses the algorithm suggested in [5] for $\rho \in (-1, 1)$ which also note the properties of this algorithm. If $\rho = -1$, we have a way to compute the probability exactly. In this case, $X = 1/Y$ almost surely, so in the following manner we compute the required probability:

\[
P(X + \frac{1}{X} > x) = P\left(X > \frac{x + \sqrt{x^2 - 2}}{2}\right) + P\left(X < \frac{x - \sqrt{x^2 - 2}}{2}\right)
\]
\[
= P\left(\log X > \log\left(\frac{x + \sqrt{x^2 - 2}}{2}\right)\right) + P\left(\log X < \log\left(\frac{x - \sqrt{x^2 - 2}}{2}\right)\right)
\]
\[
= \Phi\left(\log\left(\frac{x + \sqrt{x^2 - 2}}{2}\right)\right) + \Phi\left(\log\left(\frac{x - \sqrt{x^2 - 2}}{2}\right)\right)
\]

Patterns in the results

For judging the quality of the asymptotic approximation, we focus on the simulation estimate $P(X + Y > x)$ and not the threshold $x$, since a change of distribution may imply a change in how rare is a particular threshold crossing. So, when comparing the quality of the asymptotic approximation across different models, it makes more sense to focus on the value of $P(X + Y > x)$, rather than the particular threshold $x$. When $\rho = -1$, exact calculations suggest that the ap-
approximation is extremely good even when the actual probability $P(X + Y > x)$ is of the order of $10^{-2}$. As expected, the asymptotic approximation improves as a function of increasing threshold. When $\rho \in (-1, 1)$, we rely on the simulation estimate as a surrogate for the exact tail probability and compare it with the theoretical approximations.

The results indicate that the closer $\rho$ is to $-1$, the better the approximation. For $\rho = -1$, the approximation is good for events with probability of the order of $10^{-2}$ and to achieve comparable precision in the relative error when $\rho = 0$, the event has to be much rarer and have a probability of the order of $10^{-10}$. For $\rho = 0.9$, the results for different thresholds did not show any convergence pattern. This emphasizes that in practice the numerical approximations should be used with caution. Clearly for $\rho = 1$ the asymptotic approximation is not correct and $\rho = 0.9$ is expected to behave somewhat like the case when $\rho = 1$.

The tables give representative results. We first give the results for $\rho = -1$ in Table 2.1, since in this case no simulation is required. The column ‘Ratio’ in Table 2.1 is defined as

$$\text{Ratio} = \frac{\text{Actual probability}}{\text{Asymptotic approximation}}.$$  

For subsequent tables, the columns ‘Ratio’ and ‘Half-width’ are defined as

$$\text{Ratio} = \frac{\text{Simulation estimated probability}}{\text{Asymptotic approximation}}$$

$$\text{Half-width} = \text{Half-width of the 95\% confidence interval of the ratio.}$$

In each case, $10^7$ observations were used to compute the probability estimates.
Table 2.1: Simulation results to judge goodness of approximation when $\rho = -1$.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Actual probability</th>
<th>Asymptotic approximation</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0219</td>
<td>0.0213</td>
<td>1.0272</td>
</tr>
<tr>
<td>16</td>
<td>0.0056</td>
<td>0.0056</td>
<td>1.0121</td>
</tr>
<tr>
<td>24</td>
<td>0.0015</td>
<td>0.0015</td>
<td>1.0060</td>
</tr>
<tr>
<td>30</td>
<td>$6.7365 \times 10^{-4}$</td>
<td>$6.7091 \times 10^{-4}$</td>
<td>1.0041</td>
</tr>
<tr>
<td>100</td>
<td>$4.1233 \times 10^{-6}$</td>
<td>$4.1213 \times 10^{-6}$</td>
<td>1.0005</td>
</tr>
<tr>
<td>1000</td>
<td>$4.9238 \times 10^{-12}$</td>
<td>$4.9238 \times 10^{-12}$</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

2.6.2 How good is the portfolio suggestion?

Here, we consider the quality of our approximate solutions for the optimization problem. We choose the same risk model given in Example 2.3.6, because we have information about which values of $\rho$ lead to good asymptotic approximation. We resort to a naive method for analyzing the performance of our approximate solutions. For different $(a_1, a_2)$, we obtain estimates of $P(a_1X + a_2Y > x)$ through simulation. To get the estimates proceed as follows: For $a_1, a_2 > 0$

$$\begin{pmatrix}
  a_1 X \\
  a_2 Y
\end{pmatrix} = \begin{pmatrix}
  \exp[\log(a_1) + X_1] \\
  \exp[\log(a_2) + X_2]
\end{pmatrix}$$

Now,

$$\begin{pmatrix}
  Z_1 \\
  Z_2
\end{pmatrix} = \begin{pmatrix}
  \log(a_1) + X_1 \\
  \log(a_2) + X_2
\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}
  \log(a_1) \\
  \log(a_2)
\end{pmatrix}, \begin{pmatrix}
  1 & \rho \\
  \rho & 1
\end{pmatrix}\right)$$

$\rho \in [-1, 1)$

So, again we are in the framework of [5], and we use the algorithm given in their paper to estimate the rare event probabilities. When either $a_1$ or $a_2$ is equal
Table 2.2: Simulation results to judge goodness of approximation when $\rho = -0.9$.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Simulation estimated probability</th>
<th>Asymptotic approximation</th>
<th>Ratio</th>
<th>Half-width</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.3687</td>
<td>0.2719</td>
<td>1.3556</td>
<td>0.0006</td>
</tr>
<tr>
<td>5</td>
<td>0.1207</td>
<td>0.1075</td>
<td>1.1227</td>
<td>0.0012</td>
</tr>
<tr>
<td>10</td>
<td>0.0221</td>
<td>0.0213</td>
<td>1.0375</td>
<td>0.0026</td>
</tr>
<tr>
<td>20</td>
<td>0.0028</td>
<td>0.0027</td>
<td>1.0082</td>
<td>0.0064</td>
</tr>
<tr>
<td>30</td>
<td>$6.8873 \times 10^{-4}$</td>
<td>$6.7091 \times 10^{-4}$</td>
<td>1.0265</td>
<td>0.0119</td>
</tr>
<tr>
<td>40</td>
<td>$2.2134 \times 10^{-4}$</td>
<td>$2.2524 \times 10^{-4}$</td>
<td>0.9827</td>
<td>0.0183</td>
</tr>
<tr>
<td>50</td>
<td>$9.3675 \times 10^{-5}$</td>
<td>$9.1526 \times 10^{-5}$</td>
<td>1.0235</td>
<td>0.0285</td>
</tr>
</tbody>
</table>

to 0, we can compute the exact probability and hence do not need an estimate. We choose $(a_1, a_2)$ in the following way. Let $C$ be the set of all possible $(a_1, a_2)$ which satisfy the constraint of our optimization problem. First, $a_1$ is chosen from the corresponding projection of $C$ with a small grid, and then for each $a_1$, $a_2$ is determined from the constraint. Let us call this set $C^*$. For $(a_1, a_2) \in C^*$, $P(a_1X + a_2Y > x)$ is estimated through simulation and then it is observed which $(a_1, a_2)$ gives the minimum estimate of $P(a_1X + a_2Y > x)$. Let, $(\tilde{a}_1, \tilde{a}_2)$ be this pair; that is $P(\tilde{a}_1X + \tilde{a}_2Y > x) = \min_{(a_1, a_2) \in C^*} P(a_1X + a_2Y > x)$. Also, let $(a_1^*, a_2^*)$ be the approximate solution of the optimization problem as noted in Section 2.5.1. Relative error of the approximate solution is computed by comparing $P(a_1^*X + a_2^*Y > x)$ with $\min_{(a_1, a_2) \in C^*} P(a_1X + a_2Y > x)$. }

49
Table 2.3: Simulation results to judge goodness of approximation when $\rho = 0$.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Simulation estimated probability</th>
<th>Asymptotic approximation</th>
<th>Ratio</th>
<th>Half-width</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0338</td>
<td>0.0213</td>
<td>1.5844</td>
<td>0.0033</td>
</tr>
<tr>
<td>50</td>
<td>$1.0798 \times 10^{-4}$</td>
<td>$9.1526 \times 10^{-5}$</td>
<td>1.1798</td>
<td>0.0002</td>
</tr>
<tr>
<td>100</td>
<td>$4.5032 \times 10^{-6}$</td>
<td>$4.1213 \times 10^{-6}$</td>
<td>1.0927</td>
<td>0.0001</td>
</tr>
<tr>
<td>300</td>
<td>$1.2117 \times 10^{-8}$</td>
<td>$1.1718 \times 10^{-8}$</td>
<td>1.0341</td>
<td>0.0000</td>
</tr>
<tr>
<td>600</td>
<td>$1.6147 \times 10^{-10}$</td>
<td>$1.5853 \times 10^{-10}$</td>
<td>1.0185</td>
<td>0.0122</td>
</tr>
<tr>
<td>1000</td>
<td>$4.9821 \times 10^{-12}$</td>
<td>$4.9238 \times 10^{-12}$</td>
<td>1.0118</td>
<td>0.0000</td>
</tr>
<tr>
<td>2000</td>
<td>$1.9620 \times 10^{-14}$</td>
<td>$2.9310 \times 10^{-14}$</td>
<td>1.0106</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Identifying patterns

We do not have error estimates for our simulation results. One could consider bootstrapping to obtain such error estimates, but we have not done so. Despite the weaknesses of the naive procedure, the results are interesting.

We note one case with the linear constraint $2a_1 + 3a_2 = 1$. The suggested optimum portfolio based on asymptotic approximation is $(a_1^*, a_2^*) = (0.2, 0.2)$. The 3 cases where $\rho = -0.9, 0, 0.9$, are chosen, the reason being that we know from the results in earlier section that the asymptotic approximation is good in the case $\rho = -0.9$, reasonable when $\rho = 0$ and rather bad when $\rho = 0.9$. The approximate solution $(a_1^*, a_2^*)$ relies on replacing the original objective function by its asymptotic approximation, and so it is reasonable to expect different accuracies for these three values of $\rho$ and this turned out to be the case. In the cases of $\rho = -0.9$
Table 2.4: Simulation results to judge goodness of approximation when $\rho = 0.9$.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Simulation estimated probability</th>
<th>Asymptotic approximation</th>
<th>Ratio</th>
<th>Half-width</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0521</td>
<td>0.0213</td>
<td>2.4439</td>
<td>0.0088</td>
</tr>
<tr>
<td>30</td>
<td>0.0030</td>
<td>$6.7091 \times 10^{-4}$</td>
<td>4.4081</td>
<td>0.0275</td>
</tr>
<tr>
<td>50</td>
<td>$5.2652 \times 10^{-4}$</td>
<td>$9.1526 \times 10^{-5}$</td>
<td>5.7527</td>
<td>0.0759</td>
</tr>
<tr>
<td>75</td>
<td>$1.1217 \times 10^{-4}$</td>
<td>$1.5781 \times 10^{-5}$</td>
<td>7.1077</td>
<td>0.1843</td>
</tr>
<tr>
<td>100</td>
<td>$3.4333 \times 10^{-5}$</td>
<td>$4.1213 \times 10^{-6}$</td>
<td>8.3307</td>
<td>0.3642</td>
</tr>
</tbody>
</table>

and $\rho = 0$, we see that $\tilde{a}_1$ comes close to 0.2 as the threshold $x$ increases. But, in the case of $\rho = 0.9$, no pattern in the convergence of $\tilde{a}_1$ is observed which is expected because for $\rho = 1$, both the risks are actually the same random variable which implies indifference to the choice of $(a_1, a_2) \in C = \{(a_1, a_2) : 2a_1 + 3a_2 = 1\}$.

Another remark is that in the case where $\rho = 0.9$, the relative errors do not show any convergence pattern. We illustrate through an example the accuracy by comparing with an extreme case where we build the portfolio consisting of only one asset. For $\rho = 0$, and threshold $x = 10$, the extreme cases will yield probabilities $0.2441$ and $0.1360$. These risk probabilities are quite high compared that of our suggested optimal portfolio $(a^*_1, a^*_2)$ based on asymptotic approximation, which has risk probability $P(a^*_1X + a^*_2Y > x) = 1.0793 \times 10^{-4}$; also, the minimum of the simulation estimates $P(\tilde{a}_1X + \tilde{a}_2Y > x)$ is of the same order. So, the suggested portfolio $(a^*_1, a^*_2)$ is quite effective in reducing the risk and possibly close to the best one.

The following additional conclusion can be made. In the case of $\rho = -0.9$,
even when \( P(\tilde{a}_1 X + \tilde{a}_2 Y > x) \) is as big as 0.11, it is quite close to \( P(a^*_1 X + a^*_2 Y > x) \), indicating that the suggested optimal choice \((a^*_1, a^*_2)\) significantly reduces the risk probability. For \( \rho = 0 \), a comparable statement can be made when the minimum of the probability estimates is of the order of \( 10^{-2} \). However, for \( \rho = 0.9 \), the relative errors are never close to 0. Interestingly, even for \( \rho = 0.9 \), \( P(\tilde{a}_1 X + \tilde{a}_2 Y > x) \) and \( P(a^*_1 X + a^*_2 Y > x) \) are almost always of the same order. However, it should be noted at this point that even in this case of \( \rho = 0.9 \), the extreme cases where the portfolio is built on entirely one of the assets, \( P(a_1 X + a_2 Y > x) \) is of a much bigger order than \( P(\tilde{a}_1 X + \tilde{a}_2 Y > x) \). So, in this case, possibly \( P(a_1 X + a_2 Y > x) \) differs considerably from choices where \( a_1, a_2 > 0 \) and the case where either \( a_1 = 0 \) or \( a_2 = 0 \), but does not differ too much among the choices where \((a_1, a_2) \in C, a_1, a_2 > 0 \). This fact justifies the intuition as mentioned before that the case \( \rho = 0.9 \) is similar to case \( \rho = 1 \). Some of the results are noted in tables below.

Results are summarized in the tables given in Table 2.5, Table 2.6 and Table 2.7 for \( \rho = -0.9, 0, 0.9 \) respectively. In each case, the constraint was \( 2a_1 + 3a_2 = 1 \). For each fixed \( \rho \), we give

- the threshold \( x \),
- \( \tilde{a}_1 \), where \((\tilde{a}_1, \tilde{a}_2) \in C^* \) and

\[
P(\tilde{a}_1 X + \tilde{a}_2 Y > x) = \min_{(a_1, a_2) \in C^*} P(a_1 X + a_2 Y > x),
\]

- \( E1 = \min_{(a_1, a_2) \in C^*} P(a_1 X + a_2 Y > x), \)
- \( E2 = P(a^*_1 X + a^*_2 Y > x), \)
- the ‘Relative error’ = \( \frac{E2 - E1}{E1} \).

For each value of \( \rho \), \( a_1 \) is chosen with gap 0.01 from the projection of \( C^* \); that
is we considered \(a_1 = 0, 0.01, 0.02, \ldots, 0.5\). For each such \(a_1\), we used 10000 observations to obtain the estimates of the probability \(P(a_1X + a_2Y > x)\).

Table 2.5: Simulation study to judge effectiveness of approximate optimization when \(\rho = -0.9\).

<table>
<thead>
<tr>
<th>Threshold</th>
<th>(\tilde{a}_1)</th>
<th>E1</th>
<th>E2</th>
<th>Relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.13</td>
<td>0.1097</td>
<td>0.1204</td>
<td>0.0975</td>
</tr>
<tr>
<td>3</td>
<td>0.18</td>
<td>0.0067</td>
<td>0.0069</td>
<td>0.0322</td>
</tr>
<tr>
<td>5</td>
<td>0.19</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0294</td>
</tr>
<tr>
<td>10</td>
<td>0.19</td>
<td>1.0299 \times 10^{-4}</td>
<td>1.0592 \times 10^{-4}</td>
<td>0.0284</td>
</tr>
<tr>
<td>20</td>
<td>0.21</td>
<td>2.0806 \times 10^{-6}</td>
<td>2.0806 \times 10^{-6}</td>
<td>1.2213 \times 10^{-15}</td>
</tr>
</tbody>
</table>

2.7 Concluding Remarks

An important case for the study of asymptotic behavior of the sum of risks is the case where the risks are asymptotically independent, identically distributed and belong to the maximal domain of attraction of the Gumbel distribution. Many commonly occurring risk distributions fall in this category. We have provided sufficient conditions for

\[
\lim_{x \to \infty} \frac{P(X + Y > x)}{P(X > x)} = 2,
\]

and extended the conditions to cover the case where the marginal distributions are not the same and to the case where some risk distributions have lighter tail but the distribution does not belong to the maximal domain of attraction of the Gumbel. We are not able to provide necessary and sufficient conditions for this kind of asymptotic behavior which is an unresolved problem. It will be interest-
Table 2.6: Simulation study to judge effectiveness of approximate optimization when $\rho = 0$.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>$\tilde{a}_1$</th>
<th>E1</th>
<th>E2</th>
<th>Relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.03</td>
<td>0.1349</td>
<td>0.1723</td>
<td>0.2765</td>
</tr>
<tr>
<td>3</td>
<td>0.16</td>
<td>0.0093</td>
<td>0.0101</td>
<td>0.0759</td>
</tr>
<tr>
<td>5</td>
<td>0.18</td>
<td>0.0016</td>
<td>0.0017</td>
<td>0.0503</td>
</tr>
<tr>
<td>10</td>
<td>0.19</td>
<td>$1.0424 \times 10^{-4}$</td>
<td>$1.0793 \times 10^{-4}$</td>
<td>0.0354</td>
</tr>
<tr>
<td>20</td>
<td>0.20</td>
<td>$4.3888 \times 10^{-6}$</td>
<td>$4.3888 \times 10^{-6}$</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 2.7: Simulation study to judge effectiveness of approximate optimization when $\rho = 0.9$.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>$\tilde{a}_1$</th>
<th>E1</th>
<th>E2</th>
<th>Relative error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01</td>
<td>0.1360</td>
<td>0.1798</td>
<td>0.3223</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.0140</td>
<td>0.0208</td>
<td>0.4831</td>
</tr>
<tr>
<td>5</td>
<td>0.02</td>
<td>0.0033</td>
<td>0.0050</td>
<td>0.5146</td>
</tr>
<tr>
<td>10</td>
<td>0.02</td>
<td>$2.8357 \times 10^{-4}$</td>
<td>$4.9475 \times 10^{-4}$</td>
<td>0.7447</td>
</tr>
<tr>
<td>20</td>
<td>0.04</td>
<td>$1.3241 \times 10^{-6}$</td>
<td>$2.4023 \times 10^{-6}$</td>
<td>0.8142</td>
</tr>
</tbody>
</table>

Even for the cases where the asymptotic behavior is understood, nothing
is known about the rate of convergence in these cases; that is a quantitative estimate of how good the approximation $2P(X > x)$ is for the quantity $P(X + Y > x)$ for a large threshold $x$. Simulation studies indicate in certain circumstances the approximation is accurate, but in other cases its accuracy is dismal.

We have observed in the previous section that when tail probability approximation is good, the approximate solution of the optimization problem is also accurate whereas in the other cases this solution has poor accuracy. So, results on the rate of convergence would contribute to understanding the appropriateness of the approximate solutions in different scenarios.
3.1 Introduction

Suppose a retail company sells items each of which is covered by a warranty for a period $W$. So, the company estimates future warranty costs over a fixed period $[0, T]$, say the following quarter, based on historical data on sales and warranty claims. Typically, the length of the period $T$ for which we estimate the total warranty cost, is much smaller than the warranty period $W$. For example, for a car company, usually the warranty period $W$ is three years whereas $T$ is a quarter. We assume $2T < W$.

We consider two kinds of warranty policies, namely, the non-renewing free replacement warranty policy and the non-renewing pro-rata warranty policy. Under the first policy, the retail company agrees to repair or replace the item in case of a failure within the warranty period $W$. Under the second policy, the retail company refunds a fraction of the purchase price if the item fails within the warranty period $W$. The fraction depends on the lifetime of the item and is applicable to non-repairable items such as automobile tires; see [9, pages 133, 171].

The role and importance of warranty costs in the retail industry has increased considerably and a considerable amount of research estimates warranty costs; see [3, 29, 31, 35, 38, 55]. For the non-renewing free replacement policy case, under highly structured assumptions on the sales process and the times of claims, Kulkarni and Resnick [35] found a closed form expression for the
Laplace transform of the total warranty cost for a quarter, allowing computation of quantiles. Rather than attempting a closed form solution of the Laplace transform, we study approximations of the distribution of total warranty cost in a quarter under fairly modest assumptions on the distribution of the sales process of the warranted item and the nature of arrivals of warranty claims. Depending on the distribution of the cost of individual claims, we approximate the distribution of total warranty cost by a normal or a stable distribution. Computation of quantiles by our method is relatively straightforward. In the case of companies issuing non-renewing pro-rata warranties, we approximate the distribution of the total warranty cost by a normal distribution.

The advantage in approximating total warranty cost using our asymptotic results is that our method does not require strong assumptions on the sales process distribution or on the nature of arrival of claims, and hence is robust against model error. In practice, the times of sales may not fit the renewal or Poisson process models. Similar problems are faced when modeling times of claims and here also our method based on asymptotic results provides an alternative by doing away with the strict assumptions on the distribution of times of claims.

We discuss methods of estimating the parameters of the normal or stable distribution that approximates the distribution of the total warranty cost in $[0, T]$. We apply our methods to the sales and warranty claims data from a large car manufacturer for a single car model and model year.
3.1.1 Outline

The following sections are designed as follows. Section 3.1.2 reviews some notation. In Section 3.2, we discuss the case of non-renewing free replacement warranty policy. Section 3.3 discusses the case of the non-renewing pro-rata policy. Both kinds of warranty policies use the same assumption on the distribution of the sales process. In Section 3.4, we show that many common models for sales processes satisfy our assumptions. In Section 3.5, we propose a method of estimating parameters of the approximate distribution of the total warranty cost in $[0, T]$. Section 3.6 applies our methods to the sales and warranty claims data from a large car manufacturer for a single car model and a single model year. We close our discussion of this problem with some concluding remarks about the applicability of our results and possible future directions. The proofs of the main results are deferred to Section 3.8.

3.1.2 Notation

The point measure on $K \subset \mathbb{R}$ corresponding to the point $x$ is given by $\epsilon_x$, that is for any Borel set $A \subset K$,

$$\epsilon_x(A) = \begin{cases} 
1 & \text{if } x \in A, \\
0 & \text{otherwise}.
\end{cases}$$

The set of all Radon point measures on $K \subset \mathbb{R}$ is denoted $M_p(K)$. Similarly, the set of all non-negative Radon measures on $K \subset \mathbb{R}$ is denoted $M_+(K)$.

The set of right continuous functions with left limits from $[-W, T]$ to $\mathbb{R}$ is denoted $D([-W, T])$ and the set of continuous functions from $[-W, T]$ to $\mathbb{R}$ is de-
noted \( C([-W, T]) \). Endow \( D([-W, T]) \) with the Skorohod topology and \( C([-W, T]) \) with the uniform topology [7, pages 80, 121].

The set of all one dimensional regularly varying functions with exponent of variation \( \rho \) is written \( RV_\rho \) [47, page 24]. Also, we denote conditional expectation of \( Y \) conditioned on \( X \) as \( E^X[Y] \), that is \( E^X[Y] = E[Y|X] \).

For easy reference, we give a glossary of notation in Section 3.9.

### 3.2 Non-renewable free replacement warranty policy

The free replacement policy is the most widely used warranty policy and is used for items such as cars, consumer electronics, etc. Under this warranty, the retail company repairs or replaces the item in case of a failure within the warranty period \( W \). Typically, such policies are non-renewing [9, pages 131-133].

**Sales process:** If a warranty claim for an item comes in the period \([0, T]\), the item must be sold during the period \([-W, T]\). As a setting for our approximation procedure, imagine a family of models indexed by \( n \). Let \( S^n_j \) be the time of sale of the \( j \)-th item in the period \([-W, T]\). The sales process \( N^n(\cdot) \) is the point process

\[
N^n(t) = \sum_j \epsilon^n_j([-W, t]) = |\{j : -W \leq S^n_j \leq t\}|
\]

We further assume that \( N^n(\cdot) \) is a random element of \( D([-W, T]) \) [7, page 121].

We define a Gaussian process \( (N^{\infty}(t), t \in [-W, T]) \) having continuous paths, so \( N^{\infty}(\cdot) \) is also a random element of \( C([-W, T]) \) [7, page 80]. Existence of such a process can be guaranteed by the Kolmogorov continuity theorem [41, page 14]. We assume that the sales process \( N^n(\cdot) \), after suitable scaling and centering,
converges in distribution as \( n \to \infty \) to the limiting process \( N^\infty(\cdot) \) in \( D([-W,T]) \).

**Times of claims measures:** In the \( n \)-th model, let \( C_{nj}^n \) be the time of the \( i \)-th claim for the \( j \)-th item sold, where we start the clock at the time of sale \( S_{nj}^n \) of the \( j \)-th item, so that \( S_{nj}^n + C_{nj}^n \) is the actual claim time. Assume for all \( j \), the points \( \{C_{nj}^n, i = 1, 2, \cdots\} \) do not cluster. The times of claims measure for the \( j \)-th item \( M_j^n(\cdot) \) is

\[
M_j^n(A) = \sum_i \epsilon_{C_{nj}^n}(A), \quad A \in \mathcal{B}([0,W]),
\]

(3.2.1)

where \( \mathcal{B}([0,W]) \) is the set of all Borel subsets of \([0,W]\) and \( M_j^n(\cdot) \in \mathcal{M}_p([0,W]) \). Recall, \( W \) is the warranty period and so only claims in \([0,W]\) will be respected.

In the \( n \)-th model, assume the random measures \( \{M_j^n(\cdot), j \geq 1\} \) are independent and identically distributed for all \( j \). Moreover, assume the common distribution of the random measures remain the same for all \( n \). We denote the generic random measure describing claim times as \( M(\cdot) \), that is \( M(\cdot) \overset{d}= M_j^n(\cdot) \) for all \( j \) and all \( n \).

**Claim sizes:** We assume that the claim amounts are independent of the times of claims measures \( \{M_j^n(\cdot) : j \geq 1\} \) and the sales process \( N^n(\cdot) \) and claim amounts for different claims are independent and identically distributed.

We consolidate detailed assumptions in the following section.

### 3.2.1 Assumptions

1. Suppose \( \nu(\cdot) \) is a non-decreasing function in \( D([-W,T]) \) which is continuous at the points \( T - W \) and 0. The family of centered and scaled sales
processes in $[-W, T]$ converges weakly to a continuous path Gaussian process $N^\infty(\cdot)$ in $D([-W, T])$; that is

$$\sqrt{n} \left( \frac{N_n^\infty(\cdot)}{n} - \nu(\cdot) \right) \Rightarrow N^\infty(\cdot). \quad (3.2.2)$$

Denote the mean function of $N^\infty(\cdot)$ as $\theta(t) = E[N^\infty(t)]$ and the covariance function as $\gamma(s, t) = Cov[N^\infty(s), N^\infty(t)]$.

2. For each $n$, the times of claims measures $\{M^n_j(\cdot), j \geq 1\}$ corresponding to different items sold are independent and identically distributed and the distribution of $\{M^n_j(\cdot) : j \geq 1\}$ remains the same for all $n$. The random measure $M(\cdot)$ denotes a random element of $M_p([0, W])$ whose distribution is the same as the common distribution of $\{M^n_j(\cdot) : j \geq 1, n \geq 1\}$, that is $M(\cdot) \overset{d}{=} M^n_j(\cdot)$ for all $j$ and all $n$. For each $n$, the random measures $\{M^n_j(\cdot) : j \geq 1\}$ are all assumed to be independent of the sales process $N^n(\cdot)$.

3. The random measure $M(\cdot)$ is a Radon measure with no fixed atoms except possibly at $0$ and $W$, that is for $0 < x < W$, $P[M(\{x\})] = 0] = 1$.

4. We assume $M(\cdot)$ satisfies $E[M^2([0, W])] < \infty$.

5. For each $n$, the claim amounts for different claims are independent and identically distributed. The common distribution of the claim sizes does not change with $n$.

6. For each $n$, the claim amounts are independent of the times of claims measures $\{M^n_j(\cdot) : j \geq 1\}$ and the sales process $N^n(\cdot)$. 
3.2.2 Approximation of the distribution of total cost on warranty claims

In the $n$-th model, denote the total number of claims for the $j$-th item sold, that arrived in the fixed period $[0, T]$ by $R^n$:

$$R^n_j = \sum_i \epsilon_{S^*_j + C^*_j}([0, T])\epsilon_{C^*_j}([0, W]), \quad (3.2.3)$$

and the total number of claims in $[0, T]$ as $R^n$:

$$R^n = \sum_{\{j: -W \leq S^*_j \leq T\}} \sum_i \epsilon_{S^*_j + C^*_j}([0, T])\epsilon_{C^*_j}([0, W]) = \sum_{\{j: -W \leq S^*_j \leq T\}} R^n_j. \quad (3.2.4)$$

We require some notation to state the results. Let $r: [0, W] \to [0, 1]$ be a non-negative non-increasing function such that $r(0) = 1$. Recall the random measure $M(\cdot)$ defined in Assumption 2 of Section 3.2.1 and denote its expectation by $m(\cdot) = E[M(\cdot)]$. Then, $r(y)M(dy)$ is a random Radon measure on $[0, W]$ with expectation $\tilde{m}(\cdot)$, such that for all Borel sets $A$ of $[0, W]$,

$$\tilde{m}(A) = E \left[ \int_A r(y)M(dy) \right]. \quad (3.2.5)$$

Now, define for $x \in [-W, T]$,

$$\delta(x) = \begin{cases} 
\int_{[0, T-x]} r(y)M(dy), & \text{if } 0 \leq x \leq T, \\
\int_{[-x, 0]} r(y)M(dy), & \text{if } T - W < x < 0, \\
\int_{[-x, -W]} r(y)M(dy), & \text{if } -W \leq x \leq T - W.
\end{cases} \quad (3.2.6)$$

Note that $\delta(\cdot)$ is a random function, whose interpretation depends on the kind of warranty policy. In the free replacement warranty policy, where $r(\cdot) \equiv 1$, $\delta(x)$ gives the number of claims in $[0, T]$ for an item sold at time $x$, so that we get
The point-wise expectation and variance of $\delta(\cdot)$ are given by

$$f_1(x) = E[\delta(x)] = \begin{cases} 
\tilde{m}([0, T-x]), & \text{if } 0 \leq x \leq T, \\
\tilde{m}([-x, T-x]), & \text{if } T-W < x < 0, \\
\tilde{m}([-x, W]), & \text{if } -W \leq x \leq T-W,
\end{cases} \quad (3.2.7)$$

and

$$f_2(x) = Var[\delta(x)]. \quad (3.2.8)$$

Now, we define a function $\chi : D([-W,T]) \mapsto \mathbb{R}^{[0,W]}$ by

$$\chi(x)(u) = x(T-u) - x((-u)-), \quad x \in D([-W,T]). \quad (3.2.9)$$

Recall the Gaussian process $N^{\infty}(\cdot)$ in (3.2.2). The Gaussian random variable

$$\int_{[0,W]} \chi(N^{\infty})(u) \tilde{m}(du)$$

has expectation $\tilde{\mu}$ and variance $\tilde{\sigma}^2$ given by

$$\tilde{\mu} = \int_{[0,W]} E[\chi(N^{\infty})(u)] \tilde{m}(du) = \int_{[0,W]} E[\chi(N^{\infty})(u)] r(u)m(du),$$

$$\tilde{\sigma}^2 = \int_{[0,W]} \int_{[0,W]} Cov[\chi(N^{\infty})(u), \chi(N^{\infty})(v)] \tilde{m}(du)\tilde{m}(dv)$$

$$= \int_{[0,W]} \int_{[0,W]} Cov[\chi(N^{\infty})(u), \chi(N^{\infty})(v)] r(u)r(v)m(du)m(dv). \quad (3.2.10)$$

In the non-renewing free replacement policy, we choose $r(\cdot) \equiv 1$. Hence, the measure $\tilde{m}$ defined in (3.2.5) coincides with $m(\cdot) = E[M(\cdot)]$. Similar simplifications occur in the definitions of $\delta(\cdot)$, $f_1(\cdot)$ and $f_2(\cdot)$, as defined in (3.2.6), (3.2.7) and (3.2.8) respectively. The random function $\delta(\cdot)$ when $r(\cdot) \equiv 1$, is

$$\delta(x) = \begin{cases} 
M([0, T-x]), & \text{if } 0 \leq x \leq T, \\
M([-x, T-x]), & \text{if } T-W < x < 0, \\
M([-x, W]), & \text{if } -W \leq x \leq T-W,
\end{cases} \quad (3.2.11)$$
and the expectation and variance are given by

$$f_1(x) = E[\delta(x)] = \begin{cases} m([0, T - x]), & \text{if } 0 \leq x \leq T, \\ m([-x, T - x]), & \text{if } T - W < x < 0, \\ m([-x, W]), & \text{if } -W \leq x \leq T - W, \end{cases}$$  \tag{3.2.12}$$

and

$$f_2(x) = \text{Var}[\delta(x)].$$  \tag{3.2.13}$$

From now on, till the end of Section 3.2, we use $\delta(\cdot), f_1(\cdot)$ and $f_2(\cdot)$ to mean these simplified versions of them. We define two constants $c_1$ and $c_2$ as

$$c_1 = \int_{[-W,T]} f_1(x) \nu(dx), \quad c_2 = \int_{[-W,T]} f_2(x) \nu(dx),$$  \tag{3.2.14}$$

where $\nu(\cdot), f_1(\cdot)$ and $f_2(\cdot)$ are given in (3.2.2), (3.2.12) and (3.2.13) respectively. Also, since we chose $r(t) = 1$ for all $0 \leq t \leq W$, the parameters $\tilde{\mu}$ and $\tilde{\sigma}^2$ defined in (3.2.10) takes the simplified forms

$$\tilde{\mu} = \int_{[0,W]} E[\chi(N^\infty)(u)] m(du),$$

$$\tilde{\sigma}^2 = \int_{[0,W]} \int_{[0,W]} \text{Cov}[\chi(N^\infty)(u), \chi(N^\infty)(v)] m(du)m(dv).$$  \tag{3.2.15}$$

**Theorem 3.2.1.** Under Assumptions 1-4 of Section 3.2.1, the total number of claims $R^n$ is asymptotically normal; that is $\sqrt{n} \left( \frac{R^n}{n} - c_1 \right) \Rightarrow N(\tilde{\mu}, c_2 + \tilde{\sigma}^2)$, where $N(a, b)$ is the normal distribution with mean $a$ and variance $b$, $c_1$ and $c_2$ are given in (3.2.14) and $\tilde{\mu}$ and $\tilde{\sigma}^2$ are given in (3.2.15).

Let, $COST^n([0, T])$ be the total warranty cost during $[0, T]$ in the $n$-th model. Let $\{X_i\}$ be i.i.d. with common distribution $F(\cdot)$ representing claim sizes in $[0, T]$. Denote, $SUM_j = \sum_{i=1}^j X_i$ for all $j \geq 1$. Then, $COST^n([0, T]) = \sum_{i=1}^{R^n} X_i = SUM_{R^n}$. 64
The distribution $F(\cdot)$ of claim sizes is modeled as having a finite or infinite variance. Distributions having infinite variance are often assumed to have regularly varying tails [8, page 344]. When $F(\cdot)$ has infinite variance, we assume

$$\check{F}(\cdot) = 1 - F(\cdot) \in RV_{-\alpha}, \ 0 < \alpha < 2.$$  

The following theorem approximates the distribution of $COST^n([0, T])$ based on the assumption we make about the claim size distribution $F(\cdot)$.

**Theorem 3.2.2.** Under Assumptions 1-6 of Section 3.2.1, we approximate the total cost as follows:

1. Suppose the claim size distribution $F(\cdot)$ is such that $\int x^2 F(dx) < \infty$. Then, as $n \to \infty$,

$$\frac{COST^n([0, T]) - nc_1 E}{\sqrt{nV}} \Rightarrow N\left(\frac{E}{\sqrt{V}} \check{\mu}, c_1 + \frac{E^2}{V} (c_2 + \check{\sigma}^2)\right), \quad (3.2.16)$$

where $N(\cdot, \cdot)$, $c_1$, $c_2$, $\check{\mu}$ and $\check{\sigma}^2$ are the same as in Theorem 3.2.1, $E = \int x F(dx)$ and $V = \int x^2 F(dx) - \left(\int x F(dx)\right)^2$.

2. Suppose the claim size distribution $F(\cdot)$ is such that $\check{F}(\cdot) \in RV_{-\alpha}$, $1 < \alpha < 2$. Define $b(x) = \left(\frac{1}{1-F}\right)^{-x}(x)$. Then, as $n \to \infty$,

$$\frac{COST^n([0, T]) - nc_1 E}{b(n)} \Rightarrow c_1^{\frac{1}{\alpha}} Z_\alpha(1), \quad (3.2.17)$$

where $c_1$ is the same as in Theorem 3.2.1, $E = \int x F(dx)$ and $Z_\alpha(\cdot)$ is an $\alpha$-stable Lévy motion with $Z_\alpha(1)$ having characteristic function of the form

$$E[\exp(i\tau Z_\alpha(1))] = \exp \left( \int_0^\infty (e^{i\tau x} - 1 - i\tau x) x^{-\alpha-1} dx \right). \quad (3.2.18)$$

3. Suppose the claim size distribution $F(\cdot)$ is such that $\check{F}(\cdot) \in RV_{-\alpha}$, $0 < \alpha \leq 1$. Define $b(x) = \left(\frac{1}{1-F}\right)^{-x}(x)$ and $e(x) = \int_0^{b(x)} x F(dx)$. Then, as $n \to \infty$,

$$\frac{COST^n([0, T]) - nc_1^{\frac{1}{\alpha}} e(n)}{b(n)} \Rightarrow Z_\alpha(c_1) + 1_{\{\alpha=1\}} c_1 \log c_1, \quad (3.2.19)$$

65
where $c_1$ is the same as in Theorem 3.2.1 and $Z_\alpha(\cdot)$ is an $\alpha$-stable Lévy process with $Z_\alpha(c_1)$ having characteristic function of the form

$$E[\exp(i\tau Z_\alpha(c_1))]=\exp\left[c_1\left(\int_1^\infty (e^{i\tau x} - 1) x^{-\alpha-1} \, dx + \int_0^1 (e^{i\tau x} - 1 - i\tau x) x^{-\alpha-1} \, dx\right)\right]. \tag{3.2.20}$$

### 3.3 Non-renewable pro-rata warranty policy

The non-renewable pro-rata warranty policy is commonly used for consumer durables such as automobile batteries and tires [9, page 169]. Under this policy, the manufacturer pays a fraction of the cost of the item in case of failure within the warranty period $W$. The fraction depends on the lifetime of the item. So, if an item of cost $c_b$ fails after time $t$ from the date of purchase, the manufacturer pays the amount $q(t)$, where

$$q(t) = \begin{cases} c_b r(t) & \text{if } t \leq W, \\ 0 & \text{otherwise}, \end{cases} \tag{3.3.1}$$

where $r : [0, W] \to [0, 1]$ is a non-negative decreasing function with $r(0) = 1$. We call the function $r(\cdot)$ our rebate function. In many situations, the rebate function is taken to be a linear or quadratic function of the lifetime of the item; see [9, page 172].

In this section, since there is no repair or replacement, each item sold can have at most one warranty claim. So, the times of claims measure $M(\cdot)$ has the additional property that for any Borel measurable set $A \subset [-W, T]$, $M(A)$ can only assume two values, 0 or 1 and Assumption 4 of Section 3.2.1 is always satisfied.
3.3.1 Approximation of the distribution of total cost on warranty claims

As before, let $COST^n([0, T])$ be the total expenditure on warranty claims during the fixed period $[0, T]$ in the $n$-th model. Let $C^n_{j,t}$ be the lifetime of the $j$-th item sold. Then,

$$\sum_{\{j : S^n_j \in [-W, T]\}} c_b r(C^n_{j,t}) \epsilon C^n_{j,t}([0, W]) \epsilon S^n_j + c^n_{j,t}([0, T]).$$

Recall the definitions of $\tilde{m}(\cdot), \delta(\cdot), f_1(\cdot), f_2(\cdot)$ and $\chi(\cdot)$ given in (3.2.5), (3.2.6), (3.2.7), (3.2.8) and (3.2.9) respectively. In the case of a pro-rata warranty policy, the function $r(\cdot)$ in the definition of the random function $\delta(\cdot)$ given in (3.2.6) is the same as the rebate function $r(\cdot)$ defined in (3.3.1). Here the random function $\delta(x)$ is interpreted as the proportion of price spent on warranty claims for an item sold at time $x$. Note that the rebate function $r(\cdot)$ is known and the randomness of $\delta(\cdot)$ stems solely from the random measure $M(\cdot)$. We denote the mean and variance of the Gaussian random variable $\int_{-W}^T \chi(N^{\infty})(u) \tilde{m}(du)$ by $\tilde{\mu}$ and $\tilde{\sigma}^2$ respectively. The forms of $\tilde{\mu}$ and $\tilde{\sigma}^2$ are given in (3.2.10). We also define two constants $c_1$ and $c_2$ as

$$c_1 = \int_{[-W,T]} f_1(x) \nu(dx),\quad c_2 = \int_{[-W,T]} f_2(x) \nu(dx),$$

(3.3.2)

where $\nu(\cdot), f_1(\cdot)$ and $f_2(\cdot)$ are given in (3.2.2), (3.2.7) and (3.2.8) respectively.

**Theorem 3.3.1.** In the case of a pro-rata warranty policy, under the Assumptions 1-3 of Section 3.2.1,

$$\frac{COST^n([0, T]) - nc_b c_1}{c_b \sqrt{n}} \Rightarrow \mathcal{N}(\tilde{\mu}, c_2 + \tilde{\sigma}^2)$$

where $c_b$ is the price of each warranted item, $c_1$ and $c_2$ are given in (3.3.2) and $\tilde{\mu}$ and $\tilde{\sigma}^2$ are given in (3.2.10).
3.4 Examples of sales processes consistent with our assumptions

In the earlier two sections, we have considered the free replacement warranty policy and the pro-rata warranty policy. In both cases, the assumptions on the distribution of the times of claims measure $M(\cdot)$ as given in Assumptions 3 and 4 of Section 3.2.1 are modest and a vast class of measures qualify. In comparison, the assumption on the sales process $N^n(\cdot)$ given in (3.2.2) is stricter. Here we list several sales processes satisfying (3.2.2).

Example 3.4.1. Renewal Processes

Suppose $N(\cdot)$ is a renewal process on $[0, \infty)$, where the common inter-arrival distribution has mean $\phi_1$ and variance $\phi_2$. For the $n$-th model, define the sales process $N^n(\cdot)$ as $N^n(s) = N(n(s + W))$ for $s \in [-W, T]$ and define $B(\cdot)$ to be a Brownian motion on $[0, \infty)$. Then, from (9.4) of [47, page 293] or Theorem 14.6 of [7, page 154], we get

$$\sqrt{n} \left( \frac{1}{n} N(n(s + W)) - \frac{s + W}{\alpha} \right) \Rightarrow \frac{\sqrt{\phi_2}}{\phi_1^{1/2}} B(s + W)$$

(3.4.1)

on $D([-W, T])$. Define $\nu(s) = \frac{s + W}{\alpha}$ and $N^\infty(s) = \frac{\sqrt{\phi_2}}{\phi_1^{1/2}} B(s + W)$. The homogeneous Poisson process is a special case.

Example 3.4.2. Non-homogeneous Poisson Processes

Suppose $\nu : [-W, T] \to [0, \infty)$ is a continuous strictly increasing function and $N(\cdot)$ a homogeneous Poisson process on $[0, \infty)$ with intensity 1. Now, define the sales process $N^n(\cdot)$ as $N^n(\cdot) = N(n\nu(\cdot))$ and define $B(\cdot)$ to be a Brownian motion on $[0, \infty)$. Applying (9.4) of [47, page 293] in the case of $N(\cdot)$, we get

$$\sqrt{n} \left( \frac{1}{n} N(ns) - s \right) \Rightarrow B(s)$$

(3.4.2)
on $D([0, \infty))$. Define the composition function $\psi : D([0, \infty)) \to D([-W, T])$ by $\psi(x) = x \circ \nu$, and since $\psi(\cdot)$ is continuous [60, Theorem 3.1], using the continuous mapping theorem [7, page 21] applied to (3.4.2), we get

$$
\psi \left( \sqrt{n} \left( \frac{1}{n} N(ns) - s \right) \right) \Rightarrow \psi (B(s))
$$
on $D([-W, T])$, which implies

$$
\sqrt{n} \left( \frac{1}{n} N(n\nu(s)) - \nu(s) \right) \Rightarrow B(\nu(s))
$$
on $D([-W, T])$. Define $N_2^{\infty}(\cdot) = B(\nu(\cdot))$ and Assumption 1 of Section 3.2.1 holds.

**Example 3.4.3. Doubly Stochastic Poisson Processes**

Define $\nu(\cdot)$, $B(\cdot)$ and $N(\cdot)$ as in Example 3.4.2 and let $D_0$ be the subset of non-negative non-decreasing functions of $D([-W, T])$. Assume there exists a sequence of random elements $\{\Lambda^n(\cdot)\}$ of $D_0$ independent of $N(\cdot)$ and after centering and scaling the sequence converges to a continuous Gaussian process $N_2^{\infty}(\cdot)$ in $D([-W, T])$; that is

$$
\frac{\Lambda^n(\cdot) - n\nu(\cdot)}{\sqrt{n}} \Rightarrow N_2^{\infty}(\cdot) \quad (3.4.3)
$$
on $D([-W, T])$. Now, define the sales process $N^n(\cdot) = N(\Lambda^n(\cdot))$. Using the fact that $N(\cdot)$ is independent of $\{\Lambda^n(\cdot)\}$, (3.4.2) and (3.4.3) yield [7, page 25]

$$
\begin{bmatrix}
\sqrt{n} \left( \frac{1}{n} N(n\cdot) - \cdot \right) \\
\frac{1}{\sqrt{n}} (\Lambda^n(\cdot) - n\nu(\cdot))
\end{bmatrix} \Rightarrow
\begin{bmatrix}
B(\cdot) \\
N_2^{\infty}(\cdot)
\end{bmatrix}
$$
on $D([-W, T]) \times D([-W, T])$, where $N_2^{\infty}(\cdot)$ and $B(\cdot)$ are independent of each other. Further, using Theorem 3.9 of [7, page 37] and $\Lambda^n(\cdot)/n \Rightarrow \nu(\cdot)$ we get

$$
\begin{bmatrix}
\sqrt{n} \left( \frac{1}{n} N(n\cdot) - \cdot \right) \\
\frac{1}{n} \Lambda^n(\cdot) \\
\frac{1}{\sqrt{n}} (\Lambda^n(\cdot) - n\nu(\cdot))
\end{bmatrix} \Rightarrow
\begin{bmatrix}
B(\cdot) \\
\nu(\cdot) \\
N_2^{\infty}(\cdot)
\end{bmatrix}
$$
on $D([-W, T]) \times D_0 \times D([-W, T])$ and so, from the continuous mapping theorem [7, page 21] and Theorem 3.1 of [60], we get

$$
\begin{pmatrix}
\frac{1}{\sqrt{n}}(N(\Lambda^n(\cdot)) - \Lambda^n(\cdot)) \\
\frac{1}{\sqrt{n}}(\Lambda^n(\cdot) - n\nu(\cdot))
\end{pmatrix} \Rightarrow \begin{pmatrix} B(\nu(\cdot)) \\ N_2^\nu(\cdot) \end{pmatrix}
$$
on $D([-W, T]) \times D([-W, T])$. Therefore, applying the addition functional,

$$\sqrt{n} \left( \frac{1}{n}N(\Lambda^n(\cdot)) - \nu(\cdot) \right) \Rightarrow B(\nu(\cdot)) + N_2^\nu(\cdot)$$
on $D([-W, T])$ and the processes $B(\cdot)$ and $N_2^\nu(\cdot)$ are independent of each other. With $N_2^\nu(\cdot) = B(\nu(\cdot)) + N_2^\nu(\cdot)$, this model satisfies Assumption 1 of Section 3.2.1.

Assumption (3.4.3) is modest and Examples 3.4.1 or 3.4.2 satisfy (3.4.3).

### 3.5 Estimation procedure

An important estimation question is the choice of $n$. We interpret $n$ as a measure of the volume of sales of the warranted item. So, $n$ should depend on the size of the company and the nature of the warranted item. For example, we would expect larger $n$ for an ordinary car model than a luxury car model. We assume that for the time period we are considering, say $[−W, T]$, $n$ does not change. The non-stationarity of the sales process $N^n(\cdot)$ in this period is captured by the functions $\nu(\cdot), \theta(\cdot)$ and $\gamma(\cdot, \cdot)$ given in Assumption 1 of Section 3.2.1. If we are ambitious enough to predict the warranty cost on some time period further in future, say $[T, 2T]$, we will assume that $n$ does not change for the entire time period $[−W, 2T]$. Thus, we assume $n$ does not change for the entire time period we consider. Since we interpret $n$ as a measure of the sales volume, and $n$ does
not change for the entire time period, we choose total sales in our observed sales
data, say total sales in the time period \([-W, 0]\), for \(n\).

We discuss estimation methods for both the non-renewing free replacement
warranty policy and the non-renewing pro-rata warranty policy.

3.5.1 Free replacement policy

Which version of Theorem 3.2.2 should we apply: (1), (2) or (3)? The answer
depends on the data of claim sizes. We assumed claim sizes are i.i.d. with
common distribution function \(F(\cdot)\). A diagnostic for determining whether data
comes from a heavy-tailed distribution is the QQ plot [47, page 97]. If \(\bar{F}(\cdot) = 1 - F(\cdot) \in RV_{-\alpha}\) for some \(\alpha > 0\), we expect the QQ plot to be a straight line
with slope \(\frac{1}{\alpha}\). If we decide \(\bar{F}(\cdot) \in RV_{-\alpha}\), we estimate \(\alpha\) using one of the various
estimators of \(\alpha\) available in the literature [47, Chapter 4]. Depending on the
value of our estimate of \(\alpha\), we determine which version of Theorem 3.2.2 to use.
If our analysis yields that \(\bar{F}(\cdot) \notin RV_{-\alpha}\), we verify that \(F(\cdot)\) has finite variance and
use version (1) of Theorem 3.2.2.

For versions (1), (2) and (3) of Theorem 3.2.2, the limit relations in (3.2.16),
(3.2.17) and (3.2.19) have different sets of parameters. We proceed case by case
to discuss how we estimate parameters in each case.

Estimation of the parameters when using Theorem 3.2.2, version (1)

We estimate the six parameters given in (3.2.16): \(c_1, c_2, \bar{\mu}, \bar{\sigma}^2, E\) and \(V\). We esti-
mate \(E\) by the sample mean and \(V\) by the sample variance of the claim sizes.
For the rest of the parameters, we first analyze the sales data and estimate the functions $\nu(\cdot), \theta(\cdot)$ and $\gamma(\cdot, \cdot)$, given in Assumption 1 of Section 3.2.1. We assume that we have observed sales for the period $[-W, 0]$ and have not observed sales for the period $[0, T]$.

One parametric approach for estimating $\nu(\cdot)$, which is adopted in Section 3.6, assumes that $n\nu(\cdot)$ follows the Bass model [6] in the time period where we have observed sales, say $[-W, 0]$. Since the Bass model describes the pattern of sales from the introduction of an item in the market [6], this approach gets additional justification when we have sales data of the warranted item starting from its introduction in the market. The Bass model for total sales by time $t$, $A(t)$ (adjusted for our clock, since we have sales data for the period $[-W, 0]$) is given by

$$A(t) = n \frac{1 - \exp(-C(t + W))}{1 + (C/B - 1) \exp(-C(t + W))},$$

where $n$ is the total sales in the time period of observed sales, say $[-W, 0]$. Hence, using the Bass model for $n\nu(\cdot)$, we get that $\nu(\cdot)$ must have the form $A(t)/n$ and to estimate $\nu(\cdot)$, we have to estimate the parameters $B$ and $C$. Let $\nu'(t)$ be the density of $\nu(\cdot)$ at $t$. We minimize the squared error

$$\min_{B,C} \sum_{t=-W+1}^{0} [N^n(t) - N^n(t-1) - n\nu'(t)]^2$$

to obtain estimates $(\hat{B}, \hat{C})$. Using this procedure, we fit the Bass model to our observed data on sales (say, on the time period $[-W,0]$) and then extrapolate $\nu(\cdot)$ on some future time period, say $[0,T]$, on which we have no sales data. We denote the estimate of $\nu(\cdot)$ by $\hat{\nu}(\cdot)$. Our estimation of $\nu(\cdot)$ is free from any distributional assumption on the sales process $N^n(\cdot)$.

Now, we obtain the residuals $\{r_t = n^{-1/2} (N^n(t) - N^n(t-1) - \hat{\nu}(t) + \hat{\nu}(t-1)) : t = -W + 1, -W + 2, \cdots, 0\}$. These residuals act as surrogates for $\{N^n(t) - N^n(t-1) :$
We assume that \( N_0^\infty(t) - N_0^\infty(t-1) = T^T R_t + \mathcal{SC}_t Z(t) \), where \( Z(t) \) is a stationary Gaussian process and \( \mathcal{SC}_t \) is function of \( t \) which takes only positive values. Note that this is an additional assumption we need for estimation purposes. We have not assumed \( E[Z(t)] = 0 \) or \( \text{Var}[Z(t)] = 1 \).

We first plot the time plot of \( \{ r_t : t = -W + 1, \cdots, 0 \} \). If the time plot looks stationary, we are done and assume \( T^T R_t \equiv 0 \) and \( \mathcal{SC}_t \equiv 1 \). Otherwise, we estimate \( T^T R_t \) and \( \mathcal{SC}_t \). We do moving average smoothing on \( r_t \) to get \( \hat{T}^T R_t \), which estimates the trend. We plot absolute values of \( \{ r_t - \hat{T}^T R_t \} \) and fit another moving average estimator to it to get \( \hat{SC}_t \).

We assume \( \{ j_t = (r_t - \hat{T}^T R_t)/\hat{SC}_t : t = -W + 1, \cdots, 0 \} \) act as surrogates for the stationary process \( \{ Z(t) : t = -W + 1, \cdots, 0 \} \). We estimate the sample mean \( l \), sample variance \( s^2 \) and sample autocorrelation function \( c(\cdot) \) of \( \{ j_t \} \). Hence, \( \{ \hat{\theta}(t) - \hat{\theta}(t-1) : t = -W + 1, \cdots, 0 \} \) is estimated as

\[
\hat{\theta}(t) - \hat{\theta}(t-1) = \hat{T}^T R_t + l\hat{SC}_t,
\]

and recover \( \{ \hat{\theta}(t) : t = -W + 1, \cdots, 0 \} \). Similarly, \( \{ \text{Cov}[N_0^\infty(t) - N_0^\infty(t-1), N_0^\infty(s) - N_0^\infty(s-1)] : t = -W + 1, \cdots, 0 \} \) is estimated as

\[
\hat{\text{Cov}}[N_0^\infty(t) - N_0^\infty(t-1), N_0^\infty(s) - N_0^\infty(s-1)] = \hat{SC}_t \hat{SC}_s s^2 c(t-s).
\]

From (3.5.1), \( \{ \hat{\gamma}(\cdot, \cdot) : t, s = -W + 1, \cdots, 0 \} \) can be computed.

We also require \( \{ \hat{\theta}(t), \hat{\gamma}(t, s) : t \in [0, T], s \in [-W, T] \} \). The problem in estimating \( \{ \hat{\theta}(t), \hat{\gamma}(t, s) : t \in [0, T], s \in [-W, T] \} \) is that we do not yet have estimates of \( \{ \hat{T}^T R_t, \hat{SC}_t, c(s) : t \in [0, T], s \in [W, T + W] \} \). To get estimates of
\{\hat{T}\hat{R}_t, \hat{SC}_t, c(s) : t \in [0, T], s \in [W, T + W]\}, fit a polynomial to both \{\hat{T}\hat{R}_t : t = -W + 1, \cdots , 0\} and \{\log(\hat{SC}_t) : t = -W + 1, \cdots , 0\}. We use the fitted polynomial values to estimate \{\hat{T}\hat{R}_t, \hat{SC}_t : t \in [0, T]\}. We also assume \(c(t) = 0\) if \(t > W\), since we only have data on sales from \([-W, 0]\). If we have sales data for a longer period, then it is also possible to estimate \(c(t)\) for \(t > W\). Then, using estimates of \{\hat{T}\hat{R}_t, \hat{SC}_t, c(s) : t \in [0, T], s \in [W, T + W]\} we obtain estimates of \{\hat{\theta}(t), \hat{\gamma}(t, s) : t \in [0, T], s \in [-W, T]\} following a similar procedure as the one used to obtain \{\hat{\theta}(t), \hat{\gamma}(t, s) : t, s = -W + 1, \cdots , 0\}. Thus, we complete our estimation of \{\hat{\theta}(t), \hat{\gamma}(t, s) : t, s = -W, \cdots , T\}.

Now analyze the warranty claims data to get an estimate for the distribution of the times of claims measure \(M(\cdot)\), given in Assumption 2 of Section 3.2.1. Recall that the times of claims measure in the \(n\)-th model for the \(j\)-th item sold is \(M_j^n(\cdot)\). Also, by Assumption 2 of Section 3.2.1, \(\{M_j^n(\cdot), j = 1, 2, \cdots , n\}\) are independent and identically distributed with common distribution as that of \(M(\cdot)\). For each item \(j\) in our sales data, we consider its times of claims measure \(M_j^n(\cdot)\). If an item \(j\) has no record of claims, then we assume that \(M_j^n(\cdot) = 0\). We compute \(\{M_j^n((x - 1, x]) : x = 0, 1, \cdots , W\}\) with the interpretation that for \(x = 0\), \(M_j^n((x - 1, x]) = M_j^n([0])\). From the plot of \(\{x, \frac{1}{n} \sum_{j=1}^{n} M_j^n((x - 1, x]) : x = 0, 1, \cdots , W\}\), we infer a functional form of the mean measure \(m(\cdot) = E[M(\cdot)] = E[M_j^n(\cdot)]\). Getting a functional form of \(m(\cdot)\) is useful because to compute \(\tilde{\mu}\) and \(\tilde{\sigma}^2\), given in (3.2.15), we have to integrate with respect to \(m(dx)\); see Section 3.6.2 for an example. We denote the estimate of \(m(\cdot)\) by \(\hat{m}(\cdot)\).

Recall the definition of the parameters \(c_1\) and \(c_2\) given in (3.2.14). To estimate \(c_1\) and \(c_2\), we need to estimate first the functions \(\{\hat{f}_1(x) : x \in [-W, T]\}\) and \(\{\hat{f}_2(x) : x \in [-W, T]\}\). Actually, we estimate \(\{\hat{f}_1(x) : x = -W, -W + 1, \cdots , T\}\) and \(\{\hat{f}_2(x) : x = -W, -W + 1, \cdots , T\}\)
\[ -W, -W + 1, \cdots, T, \] and get estimates \( \hat{c}_1 = \int_{-W}^{T} \hat{f}_1(x) \hat{\nu}(dx) \) and \( \hat{c}_2 = \int_{-W}^{T} \hat{f}_2(x) \hat{\nu}(dx) \) using the trapezoid method of integration. We estimate \( \{ \hat{f}_1(x) : x = -W, \cdots, T \} \) and \( \{ \hat{f}_2(x) : x = -W, \cdots, T \} \) as

\[
\hat{f}_1(x) = \begin{cases} 
\hat{m}([0, T - x]), & \text{if } 0 \leq x \leq T, \\
\hat{m}([-x, T - x]), & \text{if } T - W < x < 0, \\
\hat{m}([-x, W]), & \text{if } -W \leq x \leq T - W,
\end{cases}
\]

(3.5.2)

and

\[
\hat{f}_2(x) = \begin{cases} 
\frac{1}{n} \sum_{j=1}^{n} [M_j^u((0, T - x))]^2 - [\hat{f}_1(x)]^2 & \text{if } 0 \leq x \leq T, \\
\frac{1}{n} \sum_{j=1}^{n} [M_j^u([-x, T - x])]^2 - [\hat{f}_1(x)]^2 & \text{if } T - W < x < 0, \\
\frac{1}{n} \sum_{j=1}^{n} [M_j^u([-x, W])]^2 - [\hat{f}_1(x)]^2 & \text{if } -W \leq x \leq T - W,
\end{cases}
\]

(3.5.3)

where \( n \) is the total number of items sold and \( M_j^u(\cdot) \) is the times of claims measure for the \( j \)-th item sold in the \( n \)-th model.

Now, we are left with the estimation of \( \hat{\mu} \) and \( \hat{\sigma}^2 \), given in (3.2.15). To estimate \( \hat{\mu} \) and \( \hat{\sigma}^2 \), first we must estimate \( \{ E[\chi(N^\infty)(u)] : u \in [0, W] \} \) and \( \{ \text{Cov}[\chi(N^\infty)(u), \chi(N^\infty)(v)] : u, v \in [0, W] \} \), where \( N^\infty(\cdot) \) is given in (3.2.2) and \( \chi(\cdot) \) is defined in (3.2.9). We estimate \( \{ E[\chi(N^\infty)(u)] : u = 0, 1, \cdots, W \} \) and \( \{ \text{Cov}[\chi(N^\infty)(u), \chi(N^\infty)(v)] : u, v = 0, 1, \cdots, W \} \) from the estimates of \( \{ \hat{\theta}(t), \hat{\gamma}(t, s) : t, s = -W + 1, \cdots, T \} \) as

\[
\hat{E}[\chi(N^\infty)(u)] = \hat{\theta}(T - u) - \hat{\theta}(u),
\]

and

\[
\hat{\text{Cov}}[\chi(N^\infty(u)), \chi(N^\infty(v))] = \hat{\gamma}(T - u, T - v) + \hat{\gamma}(-u, -v) - \hat{\gamma}(T - u, -v) - \hat{\gamma}(T - v, -u),
\]

where \( \{ \hat{\theta}(t), \hat{\gamma}(t, s) : t, s = -W + 1, \cdots, T \} \) are estimates of \( \{ \theta(t), \gamma(t, s) : t, s = -W + 1, \cdots, T \} \) obtained while analyzing the sales process. The definitions of
the functions $\theta(\cdot)$ and $\gamma(\cdot, \cdot)$ can be found in Assumption 1 of Section 3.2.1. Now, we integrate by the trapezoid method to obtain the estimated mean

$$\hat{\mu} = \int_{[0,W]} \hat{E}[\chi(N^\infty)(u)] \hat{m}(du)$$

and the estimated variance

$$\hat{\sigma}^2 = \int_{[0,W]} \int_{[0,W]} \hat{Cov}[\chi(N^\infty)(u), \chi(N^\infty)(v)] \hat{m}(du) \hat{m}(dv).$$

This method of estimation is applied to the sales and claims data of a car manufacturer for a specific model and model year in Section 3.6.

**Estimation of the parameters when using Theorem 3.2.2, version (2)**

We estimate the parameters $c_1, E, \alpha, b(n)$ and the parameters of the stable distribution of $Z_\alpha(1)$, where $Z_\alpha(1)$ is given in (3.2.17). Estimate $c_1$ and $E$ in the same manner as described in Section 3.5.1. We estimate $\alpha$ by one of its estimators [47, Chapter 4], say the QQ-estimator. There are two ways to estimate $b(n)$:

1. Use the $(1 - \frac{1}{n})$-th quantile of the i.i.d. data on claim sizes as $b(n)$; or
2. Assume the claim size distribution $F(\cdot)$ is close to Pareto and use $n^{1/\alpha}$ as an estimate of $b(n)$.

We adopt the second method of estimating $b(n)$ when analyzing data in Section 3.6.

For the stable distribution of $Z_\alpha(1)$, we follow the parameterization of [56, page 5]. From (3.2.18), we get that the parameters of the distribution of $Z_\alpha(1)$ are
Obtaining an estimate of $\sigma$ from our estimate of $\alpha$ is a simple numerical procedure.

Estimation of the parameters when using Theorem 3.2.2, version (3)

Estimate $\alpha$, say using the QQ estimator. Depending on whether $0 < \alpha < 1$ or $\alpha = 1$, our estimators will be different, but in both cases, we have to estimate the same set of parameters: $c_1, e(n), b(n)$ and the parameters of the stable distribution of $Z_\alpha(c_1)$, where $Z_\alpha(c_1)$ is given in (3.2.19). Estimate $c_1$ using the same procedure discussed in Section 3.5.1.

When $0 < \alpha < 1$, we assume that the claim size distribution $F(\cdot)$ is quite close to Pareto and hence use $n^{1/\alpha}$ as an estimate of $b(n)$ and $\frac{\alpha}{1-\alpha} \left( n^{(1-\alpha)/\alpha} - 1 \right)$ as an estimate of $e(n)$. For the stable distribution of $Z_\alpha(c_1)$, we follow the parameterization of [56, page 5]. From (3.2.20), we get that the parameters of the distribution of $Z_\alpha(c_1)$ are [56, page 170]:

$$
\mu = -\frac{c_1\alpha}{1-\alpha}, \quad \sigma = \left( c_1 \Gamma(1-\alpha) \cos\left(\frac{\pi\alpha}{2}\right) \right)^{\frac{1}{\alpha}}, \quad \beta = 1.
$$

Computing estimates of $\mu$ and $\sigma$ using our estimates of $\alpha$ and $c_1$ is routine.

If $\alpha = 1$, we assume again that the claim size distribution $F(\cdot)$ is quite close to Pareto and hence use $n$ as an estimate of $b(n)$ and $\log n$ as an estimate of $e(n)$. For the stable distribution of $Z_\alpha(c_1)$, we follow the parameterization of [56, page 5]. From (3.2.20), we get that the parameters of the distribution of $Z_\alpha(c_1)$ are [56, page 166]:

$$
\mu = 0, \quad \sigma = \left( -\frac{\Gamma(2-\alpha)}{\alpha-1} \cos\left(\frac{\pi\alpha}{2}\right) \right)^{\frac{1}{\alpha}}, \quad \beta = 1.
$$
\[ \mu = c_1 \int_0^\infty [\sin z - z 1_{\{z \leq 1\}}] z^{-2} \, dz, \quad \sigma = \frac{c_1 \pi}{2}, \quad \beta = 1. \]

Computing estimates of \( \mu \) and \( \sigma \) using our estimate of \( c_1 \) is then routine.

### 3.5.2 Pro-rata policy

The estimation method in this case is mostly similar to the one described in Section 3.5.1. We need to estimate four parameters given in Theorem 3.3.1: \( c_1, c_2, \bar{\mu} \) and \( \bar{\sigma}^2 \).

First, observe that in this case, we do not need any data on claim sizes. Given the times of claims measures \( \{M_j(\cdot) : j = 1, 2, \ldots \} \), the claim sizes are determined by the function \( q(\cdot) \) given in (3.3.1).

We analyze the sales process in the same manner as described in Section 3.5.1. Thus, we obtain estimates of the mean and covariance functions of \( \{ \chi(N^\infty)(u) : u = 0, 1, \ldots, W \} \) given by \( \{ E[\chi(N^\infty)(u)] : u = 0, \ldots, W \} \) and \( \{ \text{Cov}[\chi(N^\infty)(u), \chi(N^\infty)(v)) : u, v = 0, \ldots, W \} \). We also estimate the mean times of claims measure \( m(\cdot) = E[M(\cdot)] \) following the same methods described in Section 3.5.1. We denote the estimate of \( m(\cdot) \) by \( \hat{m}(\cdot) \).

Now, recall the parameters \( \bar{\mu} \) and \( \bar{\sigma}^2 \) given in (3.2.10) and the rebate function \( r(\cdot) \) defined in (3.3.1). Following (3.2.10), we estimate \( \bar{\mu} \) and \( \bar{\sigma}^2 \) as

\[
\hat{\mu} = \int_{[0,W]} E[\chi(N^\infty)(u)] \hat{m}(du) = \int_{[0,W]} E[\chi(N^\infty)(u)] r(u) \hat{m}(du)
\]

and

\[
\hat{\sigma}^2 = \int_{[0,W]} \int_{[0,W]} \text{Cov}[\chi(N^\infty)(u), \chi(N^\infty)(v)] \hat{m}(du) \hat{m}(dv)
\]
\[ \int_{[0,W]} \int_{[0,W]} \hat{\text{Cov}} [\chi (N^\infty)(u), \chi (N^\infty)(v)] r(u)r(v)\hat{m}(du)\hat{m}(dv). \]

Now, recall the parameters \( c_1 \) and \( c_2 \) given in (3.3.2). To compute \( c_1 \) and \( c_2 \), we first have to estimate \( \{f_1(x) : x = -W, \cdots, T\} \) and \( \{f_2(x) : x = -W, \cdots, T\} \), where the functions \( f_1(\cdot) \) and \( f_2(\cdot) \) are defined in (3.2.7) and (3.2.8). We estimate \( \{\hat{f}_1(x) : x = -W, \cdots, T\} \) and \( \{\hat{f}_2(x) : x = -W, \cdots, T\} \) as:

\[
\hat{f}_1(x) = \begin{cases} \int_{[0,T-x]} r(y)\hat{m}(dy), & \text{if } 0 \leq x \leq T, \\ \int_{[-x,T-x]} r(y)\hat{m}(dy), & \text{if } T-W < x < 0, \\ \int_{[-x,T]} r(y)\hat{m}(dy), & \text{if } -W \leq x \leq T-W, \end{cases}
\]

and

\[
\hat{f}_2(x) = \begin{cases} \frac{1}{n} \sum_{j=1}^{n} \left[ \int_{[0,T-x]} r(y)M^j_n(dy) \right] - \left[ \hat{f}_1(x) \right]^2, & \text{if } 0 \leq x \leq T, \\ \frac{1}{n} \sum_{j=1}^{n} \left[ \int_{[-x,T-x]} r(y)M^j_n(dy) \right] - \left[ \hat{f}_1(x) \right]^2, & \text{if } T-W < x < 0, \\ \frac{1}{n} \sum_{j=1}^{n} \left[ \int_{[-x,W]} r(y)M^j_n(dy) \right] - \left[ \hat{f}_1(x) \right]^2, & \text{if } -W \leq x \leq T-W, \end{cases}
\]

where \( r(\cdot) \) is the rebate function given in (3.3.1), \( n \) is the total number of items sold and \( M^j_n(\cdot) \) is the times of claims measure associated with the \( j \)-th item sold. Note that if \( r(\cdot) \equiv 1 \), the estimate of \( \{f_1(x) : x = -W, \cdots, T\} \) and \( \{f_2(x) : x = -W, \cdots, T\} \) is the same as in (3.5.2) and (3.5.3). Now, we integrate by the trapezoid method to estimate \( \hat{\varepsilon}_1 = \int_{-W}^{T} \hat{f}_1(x)\hat{v}(dx) \) and \( \hat{\varepsilon}_2 = \int_{-W}^{T} \hat{f}_2(x)\hat{v}(dx) \), where \( \hat{v}(\cdot) \) is an estimate of \( v(\cdot) \) obtained from the analysis of the sales process. The definition of \( v(\cdot) \) is given in (3.2.2).

### 3.6 Computational example

We applied our methods to automobile sales and warranty claims data from a large car manufacturer for a single car model and model year. The company
warranted each car sold for three years; that is $W = 1096$ days. The period for which we are estimating the cost is taken to be a quarter; that is $T = 91$ days. This data is the same as the one used in [35], but we do not assume the sales process or the times of claims measure $M(\cdot)$ to be Poisson.

Which version of Theorem 3.2.2 should we use? To answer this, we analyze the data on claim sizes.

### 3.6.1 Analysis of the claim size distribution

The data consists of the vehicle id which identifies the car, the date on which a car comes with some claim, the claim id which is unique for each (car, claim) pair and the amount of such a claim.

From the data, a car on a particular day could come with multiple claims. However, from our definition in (3.2.1) of the times of claims measure $M^n_j(\cdot)$ associated with the $j$-th item sold, $M^n_j(\cdot)$ is a point measure consisting of random points $\{C^n_{ji}, i = 1, 2, \cdots\}$. So, to be consistent with our modeling, for each pair (vehicle id, date), we add the costs of all the claims associated with it, that is if a car with vehicle id $V$ comes with $p$ claims on a particular date $D$, which cost $X_1, X_2, \cdots, X_p$ respectively, then we assume that the car $V$ has arrived on date $D$ with a single claim of size $X_1 + X_2 + \cdots + X_p$. Thus, we associate the claim size $(X_1 + X_2 + \cdots + X_p)$ to the (vehicle id, date) pair $(V, D)$. Our processing of the data on claim sizes differs from that of [35].

We tabulate the estimated mean, variance and quartiles of the claim size distribution in Table 3.1. Since the estimated third quartile is smaller than the
Table 3.1: Summary statistics for claim size data.

<table>
<thead>
<tr>
<th>Mean</th>
<th>Variance</th>
<th>First quartile</th>
<th>Median</th>
<th>Third quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>47.53</td>
<td>18273.14</td>
<td>7.50</td>
<td>15.91</td>
<td>42.79</td>
</tr>
</tbody>
</table>

Figure 3.1: Density and QQ plot of the claim size distribution (k = 5000, n
= 49323), Left - Density plot, Right - QQ plot.

mean, we expect power-like tails of the distribution of claim sizes. The density
plot of the claim size distribution and the QQ plot [47, page 97] are shown in
Figure 3.1. We use the QQ estimator [47, page 97] to obtain an estimate of \( \hat{\alpha} = 1.52 \). This suggests using version (2) of Theorem 3.2.2.

However, the density plot shown in Figure 3.1 almost vanishes after the
threshold 500, which suggests that there are few data points which are rela-
tively very large compared to the rest and they are heavily influencing the esti-
mate of \( \alpha \). However, there are 459 data points which are bigger than 500. So, on
one hand, we cannot discard the claim sizes which are bigger than 500 as out-
Table 3.2: Summary statistics for claim size data of size less than 500.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Variance</td>
<td>First quartile</td>
<td>Median</td>
<td>Third quartile</td>
</tr>
<tr>
<td>37.38</td>
<td>3464.91</td>
<td>7.41</td>
<td>15.73</td>
<td>41.41</td>
</tr>
</tbody>
</table>

Figure 3.2: Density and QQ plot of the distribution of claim sizes which are less than 500 (k = 5000, n = 48864), Left - Density plot, Right - QQ plot.

...liers, while on the other hand, a very small proportion of the data (459/49323 = 0.0093) is influencing the summary statistic, the QQ plot and the QQ estimate of α.

We redo the analysis for all the claim sizes which are less than 500. For this case, the summary statistics are tabulated in Table 3.2 and the density plot and the QQ plot [47, page 97] are shown in Figure 3.2. Although the data still seems to have a power-like tail, our estimate of α using the QQ estimator in this case is \( \hat{\alpha} = 2.44 \), which, to our dismay, suggests using version (1) of Theorem 3.2.2.
For comparison, we use both the versions (1) and (2) of Theorem 3.2.2 and compare the quantiles of the total warranty cost obtained from the two approximations to check robustness of our asymptotic approximation against model error.

3.6.2 Analysis of the distribution of the times of claims measure $M(\cdot)$

The sales data is needed to compute the times of claims measures $\{M^n_j(\cdot) : j = 1, 2, \cdots, n\}$. The sales data consists of the vehicle id which identifies the car and the date on which it was sold and is a record of 34807 cars sold over a period of 1116 days.

To analyze the claims data, we apply the technique in Section 3.5.1. To make any estimation about the distribution of the times of claims measure $M(\cdot)$, we obtain the data on $\{M^n_j((i - 1, i] : i = 0, 1, \cdots, W) : j = 1, 2, \cdots, n\}$. Recall, $W = 1096$ days. For each vehicle id $j$, note its date of sale $S^n_j$ and the dates on which it comes with a claim. Assume that a vehicle comes with claims on $p$ dates given by $D_1 < D_2 < \cdots < D_p$. Now, we compute $C^n_{i,j} = D_i - S^n_j, i = 1, 2, \cdots, p$. Then, we construct the measure $M^n_j(\cdot)$ as $\sum_{i=1}^{p} \epsilon_{C^n_{i,j}}(\cdot)$. In some cases, we found that $C^n_{i,j} < 0$ (claim honored before the car is sold), or $C^n_{i,j} > W$ (claim honored after the warranty period). We handled this as follows: if $C^n_{i,j} < 0$, we set $C^n_{i,j} = 0$ and if $C^n_{i,j} > W$, we set $C^n_{i,j} = W$. Thus, we obtain $\{M^n_j((i - 1, i] : i = 0, 1, \cdots, W) : j = 1, 2, \cdots, n\}$.

We estimate the expected times of claims measure $m(\cdot) = E[M(\cdot)]$ in a man-
ner similar to [35]. The plots of \( \{(i, \hat{m}_1((i - 1, i)) : i = 0, 1, \ldots, W = 1096) \} \) and \( \{(i, \hat{m}_1((i - 1, i)) : i = 1, \ldots, W - 1 = 1095) \} \) are shown in Figure 3.3, where \( n = 34807 \) is the total number of cars in our sales data. Clearly, the plot of \( \{(i, \hat{m}_1((i - 1, i)) : i = 0, 1, \ldots, 1096) \} \) indicates that the measure \( m(\cdot) \) has two atoms at 0 and \( W = 1096 \). So, we plot \( \{(i, \hat{m}_1((i - 1, i)) : i = 1, \ldots, 1095) \} \) to infer the structure of the mean times of claims measure \( m(\cdot) \) in the interval \((0, W)\).

The linear appearance of \( \{(i, \hat{m}_1((i - 1, i)) : i = 1, \ldots, 1095) \} \) as shown in Figure 3.3 suggests that for \( 0 < x < 1096 \),

\[
m(dx) = (ax + b)dx.
\]

By integrating, we get for \( 1 \leq i \leq 1095 \),

\[
m(i - 1, i) = ai + b - \frac{a}{2}.
\]

From our fitted line over \( \{(i, \hat{m}_1((i - 1, i)) : i = 1, \ldots, 1095) \} \) as shown in Figure 3.3, we obtain the estimates

\[
\hat{a} = -0.8872 \times 10^{-6}, \quad \hat{b} - \frac{\hat{a}}{2} = 0.1479 \times 10^{-2}.
\]

We estimate \( m([0]) \) and \( m([W]) \) by

\[
\hat{m}_1([0]) := \frac{1}{n} \sum_{j=1}^{n} M_j^0([0]) = 0.1330, \quad \hat{m}_1([W]) := \frac{1}{n} \sum_{j=1}^{n} M_j^0([W - 1, W]) = 0.0420.
\]

Thus, we estimate the measure \( m(\cdot) \) as \( \hat{m}(dx) = (\hat{a}x + \hat{b})dx + \hat{m}_1([0])\epsilon_0(dx) + \hat{m}_1([W])\epsilon_W(dx) \).

To estimate the parameters in the limit distribution of Theorem 3.2.2, version (1), we use the estimators of \( f_1(\cdot) \) and \( f_2(\cdot) \) suggested in (3.5.2) and (3.5.3).
3.6.3 Analysis of the distribution of the sales process

We apply the technique explained in Section 3.5.1. We assume that \( n \nu(\cdot) \) follows the Bass model [6] for the sales period of 1116 days. We choose \( n \) as the total sales in those 1116 days and so, \( n = 34807 \). We use the least squares method discussed in Section 3.5.1 to obtain estimates \((\hat{B}, \hat{C}) = (4.0149 \times 10^{-4}, 1.6738 \times 10^{-2})\). The time plot of daily count of sales with fitted Bass is given in Figure 3.4.

The fit of Bass model is even better for 12-day counts of sale as shown in Figure 3.4. In case of 12-day counts, we obtain the least square estimates \((\hat{B}, \hat{C}) = (4.0279 \times 10^{-4}, 1.6740 \times 10^{-2})\), which are not too different from the estimates obtained from daily counts. This gives us confidence in our estimates \((\hat{B}, \hat{C}) = (4.0149 \times 10^{-4}, 1.6738 \times 10^{-2})\) obtained from daily counts and we use these estimates for the following estimation procedure.

Recall from Section 3.5.1, that we have assumed that \( \{Z(t)\} \) is a stationary Gaussian process and the centered and scaled residuals \( \{j_t\} \) will act as surrogates of \( \{Z(t)\} \). We show the time plot and the normal QQ plot of \( \{j_t\} \) in Figure 3.5.
Figure 3.4: Observed sales process with fitted Bass: Left - Daily counts, Right - 12-day counts.

Figure 3.5: Time and QQ plot of \( \{ j_t \} \): Left - Time plot, Right - QQ plot.

Now, we follow the procedure described in Section 3.5.1 to estimate \( \{ \hat{\theta}(t), \hat{\gamma}(t, s) \} \).

### 3.6.4 Estimation of quantiles

First, we have to decide the time period for which we want to approximate the distribution of total warranty cost. We choose two consecutive periods of
Table 3.3: Estimated parameters $c_1, c_2, \hat{\mu}$ and $\hat{\sigma}^2$.

<table>
<thead>
<tr>
<th>Time-period</th>
<th>$\hat{c}_1$</th>
<th>$\hat{c}_2$</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\sigma}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[0, T]$</td>
<td>0.0614</td>
<td>0.0887</td>
<td>1.0210</td>
<td>1.5568</td>
</tr>
<tr>
<td>$[T, 2T]$</td>
<td>0.0540</td>
<td>0.0818</td>
<td>0.8817</td>
<td>0.9712</td>
</tr>
</tbody>
</table>

length 91 days starting from the last sales in date, that is we choose the next two quarters from the last sales date. We denote these two quarters as $[0, T]$ and $[T, 2T]$, and accordingly adjust our clock. We assume that $n$ remains the same for the entire period $[-1116, 2 \times 91]$ (recall that we have sales data for a period of 1116 days).

We compute quantiles of total warranty cost using both the stable and normal approximations and compare them.

For approximation using version (1) (normal) of Theorem 3.2.2, we follow the method described in Section 3.5.1 to obtain estimates of the six parameters: $c_1, c_2, \hat{\mu}, \hat{\sigma}^2, E$ and $V$. However, note that the parameters $c_1, c_2, \hat{\mu}$ and $\hat{\sigma}^2$ depend on the time-period we are considering, that is the estimates will be different for time-periods $[0, T]$ and $[T, 2T]$. Table 3.3 gives estimates of these parameters for the time periods $[0, T]$ and $[T, 2T]$. For both the time periods $[0, T]$ and $[T, 2T]$, we estimate the parameters $\hat{E} = 47.53$ and $\hat{V} = 18273.14$ using estimates from Table 3.1.

For approximation using version (2) (stable) of Theorem 3.2.2, we follow the method of estimation described in Section 3.5.1. We obtain estimates of $c_1$ for the time-periods $[0, T]$ and $[T, 2T]$ from Table 3.3. We estimate the parameter $\hat{E} = 47.53$ using estimates from Table 3.1. We estimate $\hat{a} = 1.52$ us-
ing the QQ estimator (k = 5000)[47, page 97] and obtain $\hat{b}(n) = n^{1/\hat{\alpha}}$. Using this estimate of $\alpha$ and (3.5.4), we estimate the parameters of the distribution of $Z_\alpha(1)$ (following the parametrization of [56], as described in Section 3.5.1) as $\hat{\mu} = 0$, $\hat{\sigma} = 1.8688$ and $\hat{\beta} = 1$. We use J. P. Nolan’s software available at http://academic2.american.edu/~jpnolan/stable/stable.html to compute the stable quantiles.

The quantiles of total warranty cost using both the approximations: version (1) and version (2) of Theorem 3.2.2, are listed in Table 3.4. Note that the quantiles of total warranty cost computed using version (1) of Theorem 3.2.2 are bigger than those computed using version (2) of Theorem 3.2.2, but the difference is not huge as one might have expected since version (2) is applicable for the heavy-tailed data whereas version (1) is applicable for the light-tailed data.

We computed the actual number of claims and the total warranty cost for the periods $[0, T]$ and $[T, 2T]$ from our data. Though we cannot test the fit of a distribution from a single observation, we do some sanity checks to decide how well the approximations work.

Start with the actual number of claims. The number of claims in $[0, T]$ is $R^n_{[0,T]} = 2352$ and the number of claims in $[T, 2T]$ is $R^n_{[T,2T]} = 1516$. Let $AF_{R^n_p}(\cdot)$ be the approximation of the distribution function of the total number of claims that arrived in period $P$ using Theorem 3.2.1. We compute $AF_{R^n_p}(R^n_p)$ for both the time periods $P = [0, T]$ and $P = [T, 2T]$. If $AF_{R^n_p}(\cdot)$ were the actual distribution function of $R^n_p$, then $AF_{R^n_p}(R^n_p)$ would be uniform on $[0, 1]$. Our computed $AF_{R^n_p}(R^n_p)$ values are $AF_{R^n_{[0,T]}}(R^n_{[0,T]}) = 0.5381$ and $AF_{R^n_{[T,2T]}}(R^n_{[T,2T]}) = 0.0029$. The probability that a $Uniform([0, 1])$ random variable take a value more extreme than a number $a$ is $2 \min\{P[U \leq a], P[U > a]\}$, where $U \sim Uniform([0, 1])$, which, for the num-
Table 3.4: Quantiles for the total cost on warranty claims.

<table>
<thead>
<tr>
<th>Time period [0, T]</th>
<th>Time period [T, 2T]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>p</strong></td>
<td><strong>p</strong>-th quantile</td>
</tr>
<tr>
<td></td>
<td>using version (1) of Theorem 3.2.2</td>
</tr>
<tr>
<td>0.50</td>
<td>110,694.91</td>
</tr>
<tr>
<td>0.75</td>
<td>119,449.01</td>
</tr>
<tr>
<td>0.80</td>
<td>121,618.18</td>
</tr>
<tr>
<td>0.85</td>
<td>124,146.62</td>
</tr>
<tr>
<td>0.90</td>
<td>127,327.97</td>
</tr>
<tr>
<td>0.95</td>
<td>132,043.22</td>
</tr>
<tr>
<td>0.99</td>
<td>140,888.23</td>
</tr>
</tbody>
</table>

The values {AF_{R_p}^n(R_p^n) : P = [0, T], [T, 2T]} are 0.9238 and 0.0058 respectively. These values suggest that {AF_{R_p}^n(·) : P = [0, T], [T, 2T]} may be reasonable fits for the distributions of {R_p^n : P = [0, T], [T, 2T]}.

Now, we compute the actual costs for the time periods [0, T] and [T, 2T], denoted by COST^w([0, T]) and COST^w([T, 2T]) respectively. Let A_1F_p(·) and A_2F_p(·) be the approximate distribution functions of the total warranty cost using versions (1) and (2) of Theorem 3.2.2 respectively for the period P. The computed values of {A_iF_p(COST^w(P)), i = 1, 2, P = [0, T], [T, 2T]} are noted in Table 3.5. For computing {A_2F_p(COST^w(P)), P = [0, T], [T, 2T]}, we used J. P. Nolan’s software available at http://academic2.american.edu/~jpnolan/stable/stable.html. If A_1F_p(·) is the actual distribution of COST^w(P), then A_1F_p(COST^w(P)) would be
Table 3.5: Values of \( \{A_iF_p(COST^n(P)), i = 1, 2, P = [0, T], [T, 2T]\} \).

<table>
<thead>
<tr>
<th>Time period ( P )</th>
<th>Actual cost in time-period ( P ) ( COST^n(P) )</th>
<th>Approximation using version (1) of Theorem 3.2.2 ( A_1F_p(COST^n(P)) )</th>
<th>Approximation using version (2) of Theorem 3.2.2 ( A_2F_p(COST^n(P)) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>([0, T])</td>
<td>148,180.60</td>
<td>0.9981</td>
<td>0.9998</td>
</tr>
<tr>
<td>([T, 2T])</td>
<td>98,992.90</td>
<td>0.5649</td>
<td>0.9983</td>
</tr>
</tbody>
</table>

uniform on \([0, 1]\). The probabilities that a \( Uniform([0, 1]) \) random variable take a value more extreme (as explained before in the previous paragraph) than \( A_1F_p(COST^n(P)) \) is greater than that of \( A_2F_p(COST^n(P)) \) for both the periods \([0, T]\) and \([T, 2T]\), which suggest that for our data on sales and warranty claims of cars, the approximation of the distribution of total warranty cost using version (1) of Theorem 3.2.2 is better than the approximation using version (2) of Theorem 3.2.2 for both the periods \([0, T]\) and \([T, 2T]\). Comparing the quantiles of \( \{A_iF_p(\cdot), i = 1, 2, P = [0, T], [T, 2T]\} \) given in Table 3.4 with the actual costs \( \{COST^n(P) : P = [0, T], [T, 2T]\} \) given in Table 3.5, we arrive at the same conclusion.

3.7 Concluding remarks

We have approximated the distribution of the total warranty claims expenses incurred in a fixed period. Our assumptions on the distribution of the sales process \( N^n(\cdot) \) and the times of claims measure \( M(\cdot) \) are mild and hence our ap-
proximation is applicable in a general context. However, we have introduced a lot of independence in our modeling. For example, we have assumed that the claim sizes are i.i.d., but in practice this may not true. Similarly, the sales process $N^n(\cdot)$ and the times of claims measures $\{M^n_j(\cdot)\}$ or the sales process $N^n(\cdot)$ and the claim sizes may be dependent. We have ignored such dependences, but allowing for dependence might lead to more realistic modeling and better approximation.

Our estimation procedure is mostly non-parametric and hence generally applicable. However, we have assumed a parametric form for $\nu(\cdot)$ using the model proposed in [6]. Estimating $\nu(\cdot)$ non-parametrically might lead to robustness against model error.

Another issue is the choice of $n$ in our approximation. We interpret $n$ as a measure of sales volume. Though total sales in our observed sales data is a natural candidate for $n$ as we have argued, it is not the only candidate. Since $n$ plays an important role in the approximation, the choice of $n$ might have a significant impact.

In Section 3.6, we have demonstrated the applicability of our method. However, a company deciding on reserves to cover warranty cost next quarter should use a more complete and carefully collected dataset.

### 3.8 Proofs

To prove the asymptotic results, first we state and prove two lemmas (recall the notations from Section 3.1.2).
Lemma 3.8.1. Under the Assumptions 1, 2 and 4 of Section 3.2.1, for \( \lambda \in \mathbb{R} \),

\[
\prod_{\{j:S_j^n \in [-W,T]\}} E^{S_j^n} \left[ \exp \left( i\lambda n^{-1/2} \left( \delta(S_j^n) - f_1(S_j^n) \right) \right) \right] = \prod_{\{j:S_j^n \in [-W,T]\}} E^{S_j^n} \left[ 1 - \lambda^2(2n)^{-1} \left( \delta(S_j^n) - f_1(S_j^n) \right)^2 \right] + o_p(1).
\]

Proof of Lemma 3.8.1. Using the fact that for all \( n \), \( \prod_{i=1}^n a_i - \prod_{i=1}^n b_i \leq \sum_{i=1}^n |a_i - b_i| \), for \( |a|, |b| \leq 1 \), we get

\[
\left| \prod_{\{j:S_j^n \in [-W,T]\}} E^{S_j^n} \left[ \exp \left( i\lambda n^{-1/2} \left( \delta(S_j^n) - f_1(S_j^n) \right) \right) \right] - \prod_{\{j:S_j^n \in [-W,T]\}} E^{S_j^n} \left[ 1 - \lambda^2(2n)^{-1} \left( \delta(S_j^n) - f_1(S_j^n) \right)^2 \right] \right| \leq \sum_{\{j:S_j^n \in [-W,T]\}} \left| E^{S_j^n} \left[ \exp \left( i\lambda n^{-1/2} \left( \delta(S_j^n) - f_1(S_j^n) \right) \right) - 1 + \lambda^2(2n)^{-1} \left( \delta(S_j^n) - f_1(S_j^n) \right)^2 \right] \right|
\]

\[
= \sum_{\{j:S_j^n \in [-W,T]\}} \left| E^{S_j^n} \left[ \exp \left( i\lambda n^{-1/2} \left( \delta(S_j^n) - f_1(S_j^n) \right) \right) - 1 - i\lambda n^{-1/2} \left( \delta(S_j^n) - f_1(S_j^n) \right) \right]^2 \right| + \lambda^2(2n)^{-1} \left( \delta(S_j^n) - f_1(S_j^n) \right)^2 \right| \]

\[
\leq \sum_{\{j:S_j^n \in [-W,T]\}} E^{S_j^n} \left[ \left( \frac{\lambda^3}{6n} \left| \delta(S_j^n) - f_1(S_j^n) \right|^3 \right) \wedge \left( \frac{\lambda^2}{n} \left( \delta(S_j^n) - f_1(S_j^n) \right)^2 \right) \right].
\]

Since for all \( x \), \( \delta(x) \) and \( f_1(x) \) are bounded by \( M([0, W]) \) and \( m([0, W]) \) respectively, the above quantity is bounded by

\[
\sum_{\{j:S_j^n \in [-W,T]\}} E \left[ \left( \frac{\lambda^3}{6n} \left( M([0, W]) + m([0, W]) \right) \right)^3 \wedge \left( \frac{\lambda^2}{n} \left( M([0, W]) + m([0, W]) \right)^2 \right) \right] \leq \frac{\lambda^2 N^n([-W,T])}{n} E \left[ \left( \frac{\lambda}{6n} \left( M([0, W]) + m([0, W]) \right) \right)^3 \wedge \left( (M([0, W]) + m([0, W]))^2 \right) \right] \]

\[
\to 0.
\]

The convergence in the last step holds since \( N^n([-W,T])/n \to \nu([-W,T]) \) by Assumption 1 of Section 3.2.1 and the quantity within the expectation converges.
to 0 using the dominated convergence theorem. To understand how we use the dominated convergence theorem, first note that the quantity inside the expectation is dominated by $(M([0, W]) + m([0, W]))^2$, which has a finite expectation by Assumption 4 of Section 3.2.1. On the other hand, the quantity inside the expectation is also dominated by \( \frac{1}{6\sqrt{n}}(M([0, W]) + m([0, W]))^3 \), which converges to 0 almost surely. Hence, using the dominated convergence theorem, we get as \( n \to \infty \),

\[
E \left[ \left( \frac{1}{6\sqrt{n}}(M([0, W]) + m([0, W])) \right)^3 \right] \to 0.
\]

\[\square\]

The sales process \( N^n(\cdot) \) is a non-decreasing process on \([-W, T]\) and hence, induces a measure on \([-W, T]\). In the following, we refer to \( N^n(\cdot) \) to mean both the sales process in \( D([-W, T]) \) and the measure it induces. It should be clear from the context what we mean by \( N^n(\cdot) \). The same rule of notation holds for the non-decreasing function \( \nu(\cdot) \) defined in (3.2.2). Now, we state the second lemma.

**Lemma 3.8.2.** Under Assumptions 1, 3 and 4 of Section 3.2.1,

1. The integral of \( f_1(\cdot) \) (defined in (3.2.7)) with respect to the centered and scaled sales process converges weakly to a Gaussian random variable, that is

\[
\int_{[-W,T]} f_1(x) \left( \frac{N^n - n\nu}{\sqrt{n}} \right) (dx) \Rightarrow \int_{[0,W]} \chi(N^\infty)(u)\tilde{m}(du),
\]

where \( \nu(\cdot) \) and \( N^\infty(\cdot) \) are given in (3.2.2), the measure \( \tilde{m}(\cdot) \) is defined in (3.2.5), the function \( \chi(\cdot) \) is defined in (3.2.9) and the Gaussian random variable \( \int_{[0,W]} \chi(N^\infty)(u)\tilde{m}(du) \) has mean \( \tilde{\mu} \) and variance \( \tilde{\sigma}^2 \) given in (3.2.10).
2. The integral of $f_2(\cdot)$ (defined in (3.2.8)) with respect to the sales process scaled by $n$ converges in probability to $c_2$ (defined in (3.3.2)), that is

$$
\frac{1}{n} \int_{[-W,T]} f_2(x) N^n(dx) \xrightarrow{P} \int_{[-W,T]} f_2(x) \nu(dx) = c_2,
$$

(3.8.4)

where $\nu(\cdot)$ is given in (3.2.2).

**Proof of Lemma 3.8.2.** (1) Step 1: We assume $2T < W$. So, $T < W - T$. First, note that, by using Fubini’s theorem, we get the following three equations

$$
\int_{[0,T]} \tilde{m}(0, T-x) \left( \frac{N^n - ny}{\sqrt{n}} \right) (dx) = \int_{[0,T]} \left( \frac{N^n - ny}{\sqrt{n}} \right) ([0, T-u]) \tilde{m}(du),
$$

(3.8.5)

$$
\int_{(T,W)} \tilde{m}([-x, T-x]) \left( \frac{N^n - ny}{\sqrt{n}} \right) (dx) = \int_{[0,T]} \left( \frac{N^n - ny}{\sqrt{n}} \right) ([0, T-u]) \tilde{m}(du) + \int_{(T,W,T)} \left( \frac{N^n - ny}{\sqrt{n}} \right) ([T-W, T-u]) \tilde{m}(du),
$$

(3.8.6)

$$
\int_{[W,T-W]} \tilde{m}([-x, W]) \left( \frac{N^n - ny}{\sqrt{n}} \right) (dx) = \int_{[W,T-W]} \left( \frac{N^n - ny}{\sqrt{n}} \right) ([T-W, T-u]) \tilde{m}(du).
$$

(3.8.7)

Using (3.8.5), (3.8.6), (3.8.7) and the definition of $\chi(\cdot)$ from (3.2.9), we get,

$$
\int_{[-W,T]} f_1(x) \left( \frac{N^n - ny}{\sqrt{n}} \right) (dx) = \int_{[0,W]} \chi \left( \frac{N^n - ny}{\sqrt{n}} \right) (u) \tilde{m}(du).
$$

(3.8.8)

Step 2: We use the continuous mapping theorem to prove our result. First, define $\xi : D([-W,T]) \mapsto \mathbb{R}$ by

$$
\xi(x) = \int_{[0,W]} \chi(x)(u) \tilde{m}(du), \quad x \in D([-W,T]),\quad (3.8.9)
$$
where $\chi(\cdot)$ is defined in (3.2.9).

We show that the continuous functions of $D([−W,T])$ are continuity points of the function $\xi : x \mapsto \int_{[0,W]} \chi(x)(u)\tilde{m}(du)$. The definition of $\xi(\cdot)$ is similar to the well-known convolution functions [25, page 143]. Suppose $x_n \to x$ in $D([−W,T])$ in the Skorohod topology and $x$ is continuous. Then, $x_n \to x$ uniformly in $[−W,T]$ [7, page 124] and

$$\left| \int_{[0,W]} \chi(x_n)(u)\tilde{m}(du) - \int_{[0,W]} \chi(x)(u)\tilde{m}(du) \right| \leq \sup_{0 \leq u \leq W} |\chi(x_n)(u) - \chi(x)(u)|\tilde{m}([0,W])$$

$$\leq \sup_{-W \leq u \leq T} 2|x_n(u) - x(u)|\tilde{m}([0,W]) \to 0,$$

as $n \to \infty$. So, the discontinuity points of $\xi(\cdot)$ are contained in $D([−W,T])\setminus C([−W,T])$.

Since the limiting process $N^\infty(\cdot) \in C([−W,T])$ and $\xi(\cdot)$ is continuous on $C([−W,T])$, by the continuous mapping theorem [7, page 21],

$$\xi \left( \frac{N^n - n\nu}{\sqrt{n}} \right) \Rightarrow \xi(N^\infty)$$

on $\mathbb{R}$, that is, using (3.8.9),

$$\int_{[0,W]} \chi \left( \frac{N^n - n\nu}{\sqrt{n}} \right)(u)\tilde{m}(du) \Rightarrow \int_{[0,W]} \chi \left( N^\infty \right)(u)\tilde{m}(du). \quad (3.8.10)$$

Hence, by (3.8.8) and (3.8.10), we have proved part (1) of Lemma 3.8.2.

(2) Using Assumption 1 of Section 3.2.1, we get $N^n(\cdot)/n \Rightarrow \nu(\cdot)$ on $D([−W,T])$. Since Skorohod convergence implies vague convergence [30], we get

$$\frac{N^n(\cdot)}{n} \Rightarrow \nu(\cdot) \quad (3.8.11)$$

in $M_1([−W,T])$ (see Section 3.1.2). By Assumption 3 of Section 3.2.1 and the definition of $\delta(\cdot)$ given in (3.2.6), the random function $\delta(\cdot)$ is almost surely continuous at all points except at most at $T−W$ and 0. Hence, using definition of $f_2(\cdot)$ in
(3.2.8), Assumption 4 of Section 3.2.1 and the dominated convergence theorem, we get that \( f_2(x) \) is discontinuous at most at two points, \( T - W \) and 0. So, \( f_2(x) \) can be written as

\[
f_2(x) = \begin{cases} 
  f_{2,c+}(x) + d_1, & \text{if } -W \leq x \leq T - W, \\
  f_{2,c+}(x) + d_2, & \text{if } T - W < x < 0, \\
  f_{2,c+}(x) + d_3, & \text{if } 0 \leq x \leq T, 
\end{cases}
\]  

(3.8.12)

where \( f_{2,c+}(x) \) is a continuous non-negative function and \( d_1, d_2, d_3 \) are three constants. Therefore, we get

\[
\frac{1}{n} \int_{[-W,T]} f_2(x) N^n(dx) = \frac{1}{n} \int_{[-W,T]} f_{2,c+}(x) N^n(dx) + \frac{d_1}{n} N^n([-W,T - W]) + \frac{d_2}{n} N^n((T - W, 0)) + \frac{d_3}{n} N^n([0, T]).
\]  

(3.8.13)

By Assumption 1 of Section 3.2.1, \( \nu(\cdot) \) is continuous at \( T - W \) and 0. Hence, using (3.8.11) we get that the last three terms on the right side of (3.8.13) converge in probability to \( d_1 \nu([-W,T - W]), d_2 \nu((T - W, 0)) \) and \( d_3 \nu([0, T]) \) respectively. Also, by (3.8.11) we get

\[
\frac{1}{n} \int_{[-W,T]} f_{2,c+}(x) N^n(dx) \overset{P}{\to} \int_{[-W,T]} f_{2,c+}(x) \nu(dx).
\]

Hence, using (3.8.12) we get

\[
\frac{1}{n} \int_{[-W,T]} f_2(x) N^n(dx) = \frac{1}{n} \int_{[-W,T]} f_{2,c+}(x) N^n(dx) + \frac{d_1}{n} N^n([-W,T - W]) + \frac{d_2}{n} N^n((T - W, 0)) + \frac{d_3}{n} N^n([0, T]) \\
\overset{P}{\to} \int_{[-W,T]} f_{2,c+}(x) \nu(dx) + \frac{d_1}{n} \nu([-W,T - W]) + \frac{d_2}{n} \nu((T - W, 0)) + \frac{d_3}{n} \nu([0, T]) \\
= \int_{[-W,T]} f_2(x) \nu(dx).
\]

\[\square\]
Proof of Theorem 3.2.1. Recall that \( R^n_j \) denotes the total number of claims in \([0, T]\) for the \( j\)-th item sold defined in (3.2.3) and \( R^n \) denotes the total number of claims in \([0, T]\) defined in (3.2.4). Using Assumption 2 of 3.2.1, we get
\[
R^n_j | S^n_j \overset{d}{=} \delta(S^n_j). \tag{3.8.14}
\]

Step 1: From Assumption 2 of Section 3.2.1, we also get that given the sales process \( N^n(\cdot) \), the random variables \( \{R^n_j : j \geq 1\} \) are independent. The characteristic function of the centered and scaled \( R^n \) given the sales process \( N^n(\cdot) \) is for \( \lambda \in \mathbb{R} \),
\[
E^{N^n(\cdot)} \left[ \exp \left( i \lambda n^{-1/2} \left( R^n - \int_{[-W,T]} f_1(x) N^n(dx) \right) \right) \right] = E^{N^n(\cdot)} \left[ \exp \left( i \lambda n^{-1/2} \sum_{\{j : S^n_j \in [-W,T]\}} (R^n_j - f_1(S^n_j)) \right) \right] = \prod_{\{j : S^n_j \in [-W,T]\}} E^{S^n_j} \left[ \exp \left( i \lambda n^{-1/2} \left( \delta(S^n_j) - f_1(S^n_j) \right) \right) \right],
\]
which, using Lemma 3.8.1, can be written as
\[
\prod_{\{j : S^n_j \in [-W,T]\}} E^{S^n_j} \left[ 1 - \lambda^2 (2n)^{-1} \left( \delta(S^n_j) - f_1(S^n_j) \right)^2 \right] + O_p(1)
\]
\[
= \prod_{\{j : S^n_j \in [-W,T]\}} \left[ 1 - \lambda^2 (2n)^{-1} f_2(S^n_j) \right] + O_p(1)
\]
\[
= \exp \left[ - \sum_{\{j : S^n_j \in [-W,T]\}} - \log \left[ 1 - \lambda^2 (2n)^{-1} f_2(S^n_j) \right] \right] + O_p(1)
\]
\[
= \exp \left[ - \int_{[-W,T]} - \log \left( 1 - \lambda^2 (2n)^{-1} f_2(x) \right) N^n(dx) \right] + O_p(1). \tag{3.8.15}
\]

Step 2: Now, since \( | - \log(1 - x) - x | \leq 2|x|^2 \) if \( |x| \leq \frac{1}{2} \), for large enough \( n \),
\[
\left| - \log \left( 1 - \frac{\lambda^2}{2n} f_2(x) \right) - \frac{\lambda^2}{2n} f_2(x) \right| \leq \frac{\lambda^4}{2n^2} f_2^2(x) \leq \frac{\lambda^4}{2n^2} \left[ E[M^2([0, W])] \right]^2 = \frac{C}{n^2}, \tag{3.8.16}
\]
where \( C = \frac{\lambda^4}{2} \left[ E[M^2([0, W])] \right]^2 \). Hence, from Lemma 3.8.2, part (2) and (3.8.16) it follows that
\[
\left( \int_{-W}^{T} \left[ - \log \left( 1 - \frac{\lambda^2}{2n} f_2(x) \right) - \frac{\lambda^2}{2n} f_2(x) \right] N^n(dx) \right) \overset{p}{\rightarrow} \left( \int_{-W}^{T} \frac{\lambda^2}{2} f_2(x) v(dx) \right).
\]
and so,
\[
\exp \left( - \int_{-W}^{T} - \log \left( 1 - \frac{\lambda^2}{2n} f_2(x) \right) N^n(dx) \right) \overset{p}{\to} \exp \left( - \int_{-W}^{T} \frac{\lambda^2}{2} f_2(x) v(dx) \right) = \exp \left( - \frac{\lambda^2 c_2^2}{2} \right). \tag{3.8.17}
\]

Therefore, using (3.8.15), (3.8.17) and the dominated convergence theorem, we get for \( \lambda \in \mathbb{R} \),
\[
E \left[ \left| E^{N_n(\cdot)} \left[ \exp \left( i \lambda n^{-1/2} \left( R^n - \int_{[-W,T]} f_1(x) N^n(dx) \right) \right) \right] - \exp \left( - \frac{\lambda^2 c_2^2}{2} \right) \right| \right] \to 0. \tag{3.8.18}
\]

Step 3: Now, we prove the joint convergence of
\[
\begin{pmatrix} X^n \\ Y^n \end{pmatrix} := \frac{1}{\sqrt{n}} \begin{pmatrix} R^n - \int_{[-W,T]} f_1(x) N^n(dx) \\ \int_{[-W,T]} f_1(x) N^n(dx) - nc \end{pmatrix}.
\]

Consider the joint characteristic function \( E \left[ \exp (i \lambda X^n + i \phi Y^n) \right] \) for \( (\lambda, \phi) \in \mathbb{R}^2 \) and note that
\[
E \left[ \exp (i \lambda X^n + i \phi Y^n) \right] = E \left[ \exp (i \lambda X^n + i \phi Y^n) - \exp \left( - \frac{\lambda^2 c_2^2}{2} + i \phi Y^n \right) \right] + E \left[ \exp \left( - \frac{\lambda^2 c_2^2}{2} + i \phi Y^n \right) \right]. \tag{3.8.19}
\]

First, we deal with the first term on the right side of (3.8.19). Note that
\[
\left| E \left[ \exp (i \lambda X^n + i \phi Y^n) - \exp \left( - \frac{\lambda^2 c_2^2}{2} + i \phi Y^n \right) \right] \right|
\leq E \left[ \left| E^{N_n(\cdot)} \left[ \exp (i \lambda X^n) \right] - \exp \left( - \frac{\lambda^2 c_2^2}{2} \right) \right| \right] \to 0,
\tag{3.8.20}
\]
where the last convergence follows from (3.8.18). For the second term on the right side of (3.8.19), observe
\[
Y^n = \left( \int_{[-W,T]} f_1(x) N^n(dx) - nc \right) / \sqrt{n} = \int_{[-W,T]} f_1(x) \left( \frac{N^n - n v}{\sqrt{n}} \right)(dx).
\]

98
Using Lemma 3.8.2, part (1), we get
\[
\lim_{n \to \infty} E \left[ \exp \left( -\frac{\lambda^2 c_2^2}{2} + i\phi Y^n \right) \right] = \exp \left( -\frac{\lambda^2 c_2^2}{2} + i\phi \tilde{\mu} - \frac{\phi^2 \tilde{\sigma}^2}{2} \right),
\]
(3.8.21)

where the parameters \( \tilde{\mu} \) and \( \tilde{\sigma}^2 \) are given in (3.2.15). Therefore, using (3.8.19), (3.8.20) and (3.8.21), we get
\[
\lim_{n \to \infty} E \left[ \exp \left( i\lambda X^n + i\phi Y^n \right) \right] = \lim_{n \to \infty} E \left[ \exp \left( -\frac{\lambda^2 c_2^2}{2} + i\phi Y^n \right) \right] = \exp \left( -\frac{\lambda^2 c_2^2}{2} + i\phi \tilde{\mu} - \frac{\phi^2 \tilde{\sigma}^2}{2} \right).
\]

So, the joint convergence holds, that is
\[
\left( X^n \right) \Rightarrow \left( X \right)
\]
on \( \mathbb{R}^2 \), where \( X \) and \( Y \) are independent, \( X \sim N(0, c_2) \) and \( Y \sim N(\tilde{\mu}, \tilde{\sigma}^2) \). Hence,
\[
\sqrt{n} \left( \frac{R^n}{n} - c_1 \right) = X^n + Y^n \Rightarrow X + Y,
\]
which gives us the required result. \( \square \)

**Proof of Theorem 3.2.2.** Let \( X_i \) denote the size of the \( i \)-th claim which arrived during the time interval \([0, T]\). Denote, \( SUM_j = \sum_{i=1}^{j} X_i \) for all \( j \geq 1 \). Then, \( COST^n([0, T]) = \sum_{i=1}^{R^n} X_i = SUM_{R^n} \).

(1) By Assumption 5 and Donsker’s theorem [7, page 146], we know
\[
\frac{SUM_{[n]} - [n]E}{\sqrt{nV}} \Rightarrow W(\cdot)
\]
on \( D([0, \infty)) \), where \( W(\cdot) \) is a standard Brownian motion. Also, from Theorem 3.2.1, we know that \( \sqrt{n} \left( \frac{R^n}{n} - c_1 \right) \) converges in distribution to a random variable \( Y \), where \( Y \sim N(\tilde{\mu}, c_2 + \tilde{\sigma}^2) \). These two facts, coupled with Assumption 6 of Section 3.2.1 gives that
\[
\left( \frac{SUM_{[n]} - [n]E}{\sqrt{nV}} \right) \Rightarrow \left( W(\cdot) \right)
\]
\[
\sqrt{n} \left( \frac{R^n}{n} - c_1 \right) \Rightarrow \left( Y \right)
\]
on $D([0, \infty)) \times \mathbb{R}$, where $W(\cdot)$ and $Y$ are independent of each other. Hence, from [7, page 37], we get
\[
\begin{pmatrix}
\sum_{[n]} - [n] E \\
\sqrt{n} \frac{R^n}{n} \\
\sqrt{n} \left( \frac{R^n}{n} - c_1 \right)
\end{pmatrix}
\Rightarrow
\begin{pmatrix}
W(\cdot) \\
c_1 \\
Y
\end{pmatrix}
\] (3.8.22)
on $D([0, \infty)) \times \mathbb{R} \times \mathbb{R}$, where each three components on the right of (3.8.22) are independent of each other. Applying Theorem 3 of [22] to (3.8.22) yields
\[
\begin{pmatrix}
\sum_{[n]} - [n] E \\
\sqrt{n} \frac{R^n}{n} \\
\sqrt{n} \left( \frac{R^n}{n} - c_1 \right)
\end{pmatrix}
\Rightarrow
\begin{pmatrix}
W(c_1(\cdot)) \\
Y
\end{pmatrix}
\] (3.8.23)
on $D([0, \infty)) \times \mathbb{R}$. Since the first component of the limit in (3.8.23) is a continuous process, we have
\[
\begin{pmatrix}
\text{COST}^{\bar{q}}(0, T) - R^n E \\
\sqrt{n} \frac{R^n}{n} \\
\sqrt{n} \left( \frac{R^n}{n} - c_1 \right)
\end{pmatrix}
\Rightarrow
\begin{pmatrix}
W(c_1) \\
Y
\end{pmatrix}
\] on $\mathbb{R} \times \mathbb{R}$ [22, Section 4]. So,
\[
\text{COST}^{\bar{q}}(0, T) - n c_1 E \\
\sqrt{n} \frac{n V}{\sqrt{V}}
\]
\[
= \text{COST}^{\bar{q}}(0, T) - R^n E \\
\sqrt{n} \frac{n V}{\sqrt{V}} + \frac{E}{\sqrt{V}} \sqrt{n} \left( \frac{R^n}{n} - c_1 \right)
\Rightarrow W(c_1) + \frac{E}{\sqrt{V}} Y
\]
on $\mathbb{R}$, which completes the proof of part (1) of Theorem 3.2.2.

(2) By Assumption 5 and a minor variant of the central limit theorem of heavy-tailed distributions [47, page 218], we know
\[
\frac{\text{SUM}_{[n]} - [n] E}{b(n)} \Rightarrow Z_\alpha(\cdot)
\] on $D([0, \infty))$, where $Z_\alpha(\cdot)$ is an $\alpha$-stable Lévy motion whose characteristic function satisfies (3.2.18). Now, applying Theorem 3 of [22] in a similar way as in the proof of (3.2.16), we get
\[
\frac{\text{SUM}_{[R^n]} - [R^n] E}{b(n)} \Rightarrow Z_\alpha(c_1(\cdot)),
\] (3.8.24)
on $D([0, \infty))$. Since the limit process in (3.8.24) is continuous in probability at time 1, we have
\[
\frac{\text{COST}^a([0, T]) - R^a E}{b(n)} \Rightarrow Z_\alpha(c_1) \overset{d}{=} c_1^{\frac{1}{a}} Z_\alpha(1) \tag{3.8.25}
\]
on $\mathbb{R}$. Observe,
\[
\frac{\text{COST}^a([0, T]) - nc_1 E}{b(n)} = \frac{\text{COST}^a([0, T]) - R^a E}{b(n)} + \frac{R^a E - nc_1 E}{b(n)}.
\tag{3.8.26}
\]
In (3.8.25), we have already found the weak limit of the first term on the right side of (3.8.26). Now, we deal with the second term on the right side of (3.8.26).
Notice, $b(\cdot) \in RV_{\frac{\alpha}{2}}$, and hence $x^{1/2}(b(x))^{-1} \in RV_{\frac{1}{2} - \frac{1}{\alpha}}$. Since $\alpha < 2$, $\frac{1}{2} - \frac{1}{\alpha} < 0$. Hence, $\lim_{n \to \infty} \sqrt{n}/b(n) = 0$. Therefore, using this fact, along with Theorem 3.2.1 and Slutsky’s theorem, we get
\[
\frac{R^a E - nc_1 E}{b(n)} = \sqrt{n} \frac{E(b(x))}{b(n)} \sqrt{n} \left( \frac{R^a}{n} - c_1 \right) \overset{p}{\to} 0,
\]
which together with (3.8.25) and (3.8.26) completes the proof of part (2) of Theorem 3.2.2.

(3) By Assumption 5 and the central limit theorem of heavy-tailed distributions [47, page 218], we know
\[
\frac{\sum_{[n]} - [n]e(n)}{b(n)} \Rightarrow Z_\alpha(\cdot)
\]
on $D([0, \infty))$, where $Z_\alpha(\cdot)$ is an $\alpha$-stable Lévy process and characteristic function of $Z_\alpha(c_1)$ satisfies (3.2.20). Now, applying Theorem 3 of [22] in a similar way as in the proof of (3.2.16), we get
\[
\frac{\sum_{[R^n]} - [R^n]e(R^n)}{b(n)} \Rightarrow Z_\alpha(c_1(\cdot)) \tag{3.8.27}
\]
on $D([0, \infty))$ and since the limit process in (3.8.27) is continuous in probability at time 1, we get
\[
\frac{\text{COST}^a([0, T]) - R^a e(R^n)}{b(n)} \Rightarrow Z_\alpha(c_1) \tag{3.8.28}
\]
on \( \mathbb{R} \). Now, notice that
\[
\frac{\text{COST}^n([0, T]) - nc_1^{\frac{1}{\alpha}} e(n)}{b(n)} = \frac{\text{COST}^n([0, T]) - R^n e(R^n)}{b(n)} + \frac{R^n e(R^n) - nc_1^{\frac{1}{\alpha}} e(n)}{b(n)}. \tag{3.8.29}
\]

We have already observed in (3.8.28) the weak convergence of the first term on the right side of (3.8.29). We now turn to the second term. To show the convergence in probability of the second term, we consider two separate cases, namely \( \alpha = 1 \) and \( \alpha < 1 \).

\( \alpha < 1 \) case: Notice \( e(x) = I(b(x)) \) where \( I(x) = \int_0^x tF(dt) \in RV_{1-\alpha} \) [47, page 36, Ex. 2.5]. Also, \( b(x) \in RV_{1-\alpha} \) and \( \lim_{x \to \infty} b(x) = \infty \). Therefore, from Proposition 2.6(iv) of [47, page 32], \( e(\cdot) \in RV_{\frac{1}{1-\alpha}} \). So, \( xe(x) \in RV_1 \). Also, from Theorem 3.2.1, we know that \( R^n/n \) converges in probability to \( c_1 \). Hence, [47, page 36]
\[
\frac{R^n e(R^n)}{ne(n)} \rightarrow c_1^{\frac{1}{\alpha}}. \tag{3.8.30}
\]

Using (3.8.30) and the fact that the sequence \( \frac{ne(n)}{b(n)} \) converges to \( \frac{a}{1-\alpha} \) and hence is bounded, we get that
\[
\frac{R^n e(R^n) - nc_1^{\frac{1}{\alpha}} e(n)}{b(n)} = \frac{ne(n)}{b(n)} \left( \frac{R^n e(R^n)}{ne(n)} - c_1^{\frac{1}{\alpha}} \right) \overset{p}{\rightarrow} 0,
\]
which, together with (3.8.28) and (3.8.29) completes the proof for the case \( \alpha < 1 \).

\( \alpha = 1 \) case: In this case, we may write the second term of the right side of (3.8.29) as
\[
\frac{R^n e(R^n) - nc_1 e(n)}{b(n)} = \frac{R^n e(R^n) - R^n e(n)}{b(n)} + \frac{R^n e(n) - nc_1 e(n)}{b(n)}
= \frac{R^n e(R^n) - e(n)}{n b(n)/n} + \frac{\sqrt{ne(n)}}{b(n)} \sqrt{n} \left( \frac{R^n}{n} - c_1 \right). \tag{3.8.31}
\]

First we deal with the first term on the right side of (3.8.31). Note that \( \tilde{F}(\cdot) \) is (-1)-varying and so, using a theorem of [16] ([45, page 30, Proposition 0.11]),
we get that \( \int_{0}^{x} \tilde{F}(y) \, dy \) is \( \Pi \)-varying with auxiliary function \( x\tilde{F}(x) \). Hence, using the fact that \( I(x) = \int_{0}^{x} \tilde{F}(y) \, dy - x\tilde{F}(x) \), it easily follows from the definition of \( \Pi \)-varying functions [45, page 27], that \( I(x) \) is also \( \Pi \)-varying with auxiliary function \( x\tilde{F}(x) \). Since, \( b(x) \in RV_1 \), \( e(x) = I(b(x)) \) is also \( \Pi \)-varying with auxiliary function \( b(x)\tilde{F}(b(x)) = b(x)/x \) [20], [47, page 38].

Using local uniform convergence of \( \Pi \)-varying functions on sets away from 0 [8, page 139, Theorem 3.1.16], and the fact \( R^n/n \overset{p}{\rightarrow} c_1 > 0 \), we get

\[
\frac{e(R^n) - e(n)}{b(n)/n} = \frac{e\left(n \cdot \frac{R^n}{n}\right) - e(n)}{b(n)/n} \overset{p}{\rightarrow} \log c_1.
\]

Hence,

\[
\frac{R^n e(R^n) - R^n e(n)}{b(n)} = \frac{R^n e(R^n) - e(n)}{n} \frac{e(n)}{b(n)/n} \overset{p}{\rightarrow} c_1 \log c_1. \tag{3.8.32}
\]

Now, we turn to the second term in (3.8.31). Note that \( b(n) \in RV_1 \) and since \( e(n) \) is \( \Pi \)-varying, \( e(n) \in RV_0 \) [8, page 128]. Therefore, \( \sqrt{n} e(n)/b(n) \in RV_{-1} \), and so, \( \lim_{n \to \infty} \frac{\sqrt{n} e(n)}{b(n)} = 0 \). Also, from Theorem 3.2.1, we know that \( \sqrt{n} \left( \frac{e(n)}{n} - c_1 \right) \) is asymptotically normal. Therefore, using Slutsky’s theorem, we get

\[
\frac{R^n e(n) - nc_1 e(n)}{b(n)} = \frac{\sqrt{n} e(n)}{b(n)} \sqrt{n} \left( \frac{R^n}{n} - c_1 \right) \overset{p}{\rightarrow} 0. \tag{3.8.33}
\]

Finally, using (3.8.31), (3.8.32) and (3.8.33) we get

\[
\frac{R^n e(R^n) - nc_1 e(n)}{b(n)} \overset{p}{\rightarrow} c_1 \log c_1,
\]

which, together with (3.8.28) and (3.8.29) completes the proof when \( \alpha = 1 \). \( \square \)

**Proof of Theorem 3.3.1.** Denote by \( COST^n_j([0, T]) \) the warranty claims expenditures which are incurred in \([0, T]\) for the \( j \)-th item, sold in the time interval \([-W, T]\). In notation,

\[
COST^n_j([0, T]) = c_{br}(C^n_{j1}) \mathcal{E}_{C^n_{j1}}([0, W]) \mathcal{E}_{\mathcal{S}^{\nu+\nu}_{j1}}([0, T]).
\]
So, $\text{COST}^n([0, T]) = \sum_{(j : S^n_j \in [-W, T])} \text{COST}^n_j([0, T])$. Using Assumption 2 of Section 3.2.1, we get

$$(c_b)^{-1} \text{COST}^n_j([0, T])|S^n_j \overset{d}{=} \delta(S^n_j).$$

Now, we consider the characteristic function of

$$\left((c_b)^{-1} \text{COST}^n([0, T]) - \int_{[-W,T]} f_1(x)N^n(dx)\right)/\sqrt{n}$$

given the sales process $N^n(\cdot)$. Note that for $\lambda \in \mathbb{R}$,

$$E^{N^n(\cdot)} \left[ \exp \left(i \lambda n^{-1/2} \left((c_b)^{-1} \text{COST}^n([0, T]) - \int_{[-W,T]} f_1(x)N^n(dx)\right)\right) \right] = E^{N^n(\cdot)} \left[ \exp \left(i \lambda n^{-1/2} \sum_{(j : S^n_j \in [-W,T])} \left((c_b)^{-1} \text{COST}^n_j([0, T]) - f_1(S^n_j)\right)\right) \right] = \prod_{(j : S^n_j \in [-W,T])} E^{S^n_j} \left[ \exp \left(i \lambda n^{-1/2} \left(\delta(S^n_j) - f_1(S^n_j)\right)\right) \right].$$

Now, we want to use the same steps as in the proof of Theorem 3.2.1. To do so, we must be able to use Lemma 3.8.1 and Lemma 3.8.2, which need Assumptions 1-4 of Section 3.2.1. By the hypothesis of Theorem 3.3.1, we have Assumptions 1-3 of Section 3.2.1. Since $M([0, W]) \leq 1$, we know that Assumption 4 of Section 3.2.1 is also satisfied. Hence, we can apply the results of Lemma 3.8.1 and Lemma 3.8.2 here, too. Now, the proof follows exactly similar steps as in the proof of Theorem 3.2.1. So, the rest of the proof is omitted. □

### 3.9 Glossary of notation for Chapter 3

$S^n_j = $ time of sale of the $j$-th item in the $n$-th model,

$N^n(t) = $ total number of sales in $[-W, t]$ in the $n$-th model $= \sum_j \epsilon_{S^n_j}([-W, t])$.  

104
\( \nu(t) = \) the centering function of \( \frac{1}{n}N^n(t) \),

\( N^\infty(\cdot) = \) the Gaussian process limit of centered and scaled \( \frac{1}{n}N^n(\cdot) \),

\( C_{j,i}^n = \) the time of the \( i \)-th claim for the \( j \)-th item sold, where we start our clock at the time of sale \( S_j^n \),

\( M_j^n(\cdot) := \sum_i \epsilon C_{j,i}^n(\cdot) = \) the times of claims measure for the \( j \)-th item sold,

\( M(\cdot) = \) the generic random measure representing times of claims for an item sold, that is \( M(\cdot) \overset{d}{=} M_j^n(\cdot) \) for all \( j \) and all \( n \).

\( m(\cdot) = E[M(\cdot)] \),

\( \hat{m}_1(\cdot) = \frac{1}{n} \sum_{j=1}^n M_j^n(\cdot) \),

\( \hat{m}(\cdot) = \) estimate of \( m(\cdot) \), maybe using a functional form,

\( R_j^n = \) total number of claims in \([0, T]\) for the \( j \)-th item sold

\[ = \sum_i \epsilon C_{j,i}^n([0, T]) \epsilon C_{j,i}^n([0, W]) \],

\( R^n = \) total number of claims in \([0, T]\) = \( \sum_{\{j:S_j^n \in [-W, T]\}} R_j^n \),

\( r(t) = \begin{cases} 
1, & \text{if we follow the non-renewing free replacement policy,} \\
\text{the fraction of the cost} & \text{if we follow the non-renewing refunded if the claim pro-rata policy,} \\
& \text{comes after time} \ t \\
& \text{from the date of sale,}
\end{cases} \)

\( \tilde{m}(\cdot) = E \left[ \int_r(\cdot) M(dy) \right] \),
\[ \delta(x) = \begin{cases} 
\int_{[0,T-x]} r(y)M(dy), & \text{if } 0 \leq x \leq T, \\
\int_{[T-x,T-x]} r(y)M(dy), & \text{if } T - W < x < 0, \\
\int_{[x]} r(y)M(dy), & \text{if } -W \leq x \leq T - W, 
\end{cases} \]

\[ f_1(x) = E[\delta(x)] = \begin{cases} 
\tilde{m}([0,T-x]), & \text{if } 0 \leq x \leq T, \\
\tilde{m}([-x,T-x]), & \text{if } T - W < x < 0, \\
\tilde{m}([-x,W]), & \text{if } -W \leq x \leq T - W, 
\end{cases} \]

\[ f_2(x) = \text{Var}[\delta(x)], \]

\[ c_1 = \int_{-W}^{T} f_1(x)\nu(dx), \]

\[ c_2 = \int_{-W}^{T} f_2(x)\nu(dx), \]

\[ E = \text{expectation of the distribution of claim sizes, if it exists}, \]

\[ V = \text{variance of the distribution of claim sizes, if it exists}, \]

\[ \chi : D([-W,T]) \to \mathbb{R}^{[0,W]} \text{ defined as} \]

\[ \chi(x)(u) = x(T-u) - x((-u)-), \]

\[ \tilde{\mu} = \int_{[0,W]} E[\chi(N^\infty)(u)]\tilde{m}(du), \]

\[ \tilde{\sigma}^2 = \int_{[0,W]} \int_{[0,W]} \text{Cov}[\chi(N^\infty)(u),\chi(N^\infty)(v)]\tilde{m}(du)\tilde{m}(dv). \]
CHAPTER 4
HIDDEN REGULAR VARIATION: DETECTION AND ESTIMATION

4.1 Introduction

Multivariate risks with Pareto-like tails are usually modeled using the theory of regular variation on cones. Let $C$ be a cone in $[0, \infty]^d$ satisfying $x \in C$ implies $tx \in C$ for $t > 0$. The topology on $[0, \infty]^d$ is homeomorphic to the Euclidean topology on $[0, 1]^d$ and whenever we consider the topology on a subset $A$ of $[0, \infty]^d$, we endow $A$ with the relative topology induced by the topology on $[0, \infty]^d$. Denote the set of all non-negative Radon measures on $C$ by $M_+(C)$. The distribution of a random vector $Z$ is regularly varying on $C$ if there exist a scaling function $g(t) \uparrow \infty$, and a non-zero Radon measure $\chi(\cdot) \in M_+(C)$ such that (1.3.1) is satisfied in $M_+(C)$. Risks with heavy tails could also be modeled by stable distributions on a general convex cone; see [14].

Suppose the distribution of a random vector $Z$ is regularly varying on the first quadrant $E := [0, \infty]^d \setminus \{(0, 0, \cdots, 0)\}$ as in (1.3.1) with limit measure $\nu(\cdot)$. It is possible for $\nu(\cdot)$ to give zero mass to a proper sub-cone $C \subseteq \mathbb{E}$; for example, we could have

$$C = \mathbb{E}^{(2)} = \mathbb{E} \setminus \bigcup_{1 \leq j_1 < j_2 < \cdots < j_{d-1} \leq d} \{x^{j_1} = 0, \cdots, x^{j_{d-1}} = 0\},$$

the first quadrant with the axes removed. If the distribution of $Z$ is also regularly varying on the subcone $C$ with scaling function $g_C(t) \uparrow \infty$ and $g(t)/g_C(t) \to \infty$, then we say the distribution of $Z$ possesses hidden regular variation (HRV) on $C$. HRV helps detect finer structure that may be ignored by regular variation on $\mathbb{E}$. We will later refine our definition of hidden regular variation for a finite
sequence of cones $\mathbb{E} \supset C_1 \supset C_2 \supset \cdots \supset C_m$.

Failure of regular variation on $\mathbb{E}$ to distinguish between independence and asymptotic independence prodded Ledford and Tawn [36, 37] to define the co-efficient of tail dependence and this idea was extended to hidden regular variation on $\mathbb{E}^{(2)}$ in [46]. See also [12, 17, 21, 27, 28, 39, 42, 43, 44, 51, 58].

Hidden regular variation provides models that possess regular variation on $\mathbb{E}$ and asymptotic independence [47, pages 323-325]. The concept has typically been considered in two dimensions using the sub-cone $\mathbb{E}^{(2)}$. It is not clear how best to extend the ideas of HRV to dimensions higher than two and one obvious remark is that how one proceeds with definitions depends on the sort of risk regions being considered.

To demonstrate what is possible in higher dimensions, in this chapter we define hidden regular variation on the sub-cones

$$\mathbb{E}^{(l)} = [0, \infty]^d \setminus \bigcup_{1 \leq j_1 < j_2 < \cdots < j_{d-l+1} \leq d} \{x^{j_1} = 0, \cdots, x^{j_{d-l+1}} = 0\}, \quad 3 \leq l \leq d,$$

of $\mathbb{E}$ and show with an example that asymptotic independence is not a necessary condition for HRV on $\mathbb{E}^{(l)}$, $3 \leq l \leq d$. Hidden regular variation on $\mathbb{E}^{(l)}$, means that the distribution of the random vector $Z$ is regularly varying on $\mathbb{E}$ as in (1.3.1) with limit measure $\nu(\cdot)$ and $\nu(\mathbb{E}^{(l-1)}) > 0$, but $\nu(\mathbb{E}^{(l)}) = 0$. Also, there is a scaling function $g_{\mathbb{E}^{(l)}}(t) \uparrow \infty$ satisfying $g(t)/g_{\mathbb{E}^{(l)}}(t) \to \infty$ which makes the distribution of $Z$ regularly varying on the cone $\mathbb{E}^{(l)}$ as in (1.3.1) with limit measure $\nu^{(l)}(\cdot)$. Later, when we define HRV on the finite sequence of cones $\mathbb{E} \supset \mathbb{E}^{(2)} \supset \cdots \supset \mathbb{E}^{(d)}$, our definition of HRV on $\mathbb{E}^{(l)}$ will be modified accordingly. We suggest exploratory methods for detecting the presence of hidden regular variation on $\mathbb{E}^{(l)}$, $2 \leq l \leq d$. The existing method of detecting hidden regular variation on $\mathbb{E}^{(2)}$ is valid only for dimension $d = 2$, but our detection methods are applicable for any finite
dimension.

If exploratory detection methods confirm that the data is consistent with the hypothesis of regular variation on a cone $E^{(l)}$ as in (1.3.1), we must estimate the limit measure $\nu^{(l)}(\cdot)$. Previous methods [27] for estimating the limit measure $\nu^{(2)}(\cdot)$ of hidden regular variation on $E^{(2)}$ have been non-parametric and ignored the semi-parametric structure of $\nu^{(2)}(\cdot)$. We offer some improvement by exploiting the semi-parametric structure of $\nu^{(2)}(\cdot)$ and estimate the parametric and non-parametric parts of $\nu^{(2)}(\cdot)$ separately.

On $E$, estimation of the limit measure of regular variation is resolved by the familiar method of the polar coordinate transformation $x \rightarrow (||x||, x/||x||)$; after this transformation, the limit measure $\nu(\cdot)$ is a product of a probability measure $S(\cdot)$ and a Pareto measure $\nu_a(\cdot), \nu_a((r, \infty)) = r^{-q}, r > 0$ [47, pages 168-179]. Trying to decompose $\nu^{(2)}(\cdot)$ in this way presents the difficulty that the decomposition gives a Pareto measure $\nu_{a^{(2)}}(\cdot)$ and a possibly infinite Radon measure [47, pages 324-339]. So we transform to a different coordinate system after which $\nu^{(2)}(\cdot)$ is a product of a Pareto measure $\nu_{a^{(2)}}(\cdot)$ and a probability measure $S^{(2)}(\cdot)$ on $\delta \mathbb{N}^{(2)} = \{x \in \mathbb{E}^{(2)} : x^{(2)} = 1\}$, where $x^{(2)}$ is the second largest component of $x$. We call the probability measure $S^{(2)}(\cdot)$ the hidden angular measure on $\mathbb{E}^{(2)}$. We suggest procedures for consistently estimating the parameter $a^{(2)}$ of the Pareto measure $\nu_{a^{(2)}}(\cdot)$ and the hidden angular measure $S^{(2)}(\cdot)$ and explain how these estimates lead to an estimate of $\nu^{(2)}(\cdot)$. If HRV on $E^{(l)}$ is present for some $3 \leq l \leq d$, there is a similar transformation of coordinates making $\nu^{(l)}(\cdot)$ a product of a Pareto measure $\nu_{a^{(l)}}(\cdot)$ and a probability measure $S^{(l)}(\cdot)$ on $\delta \mathbb{N}^{(l)} = \{x \in \mathbb{E}^{(l)} : x^{(l)} = 1\}$, where $x^{(l)}$ is the $l$-th largest component of $x$. We call this probability measure $S^{(l)}(\cdot)$ the hidden angular measure on $\mathbb{E}^{(l)}$ and employ similar estimation methods.
for $l \geq 3$ as we did for $l = 2$.

For empirical exploration of the angular or hidden angular measures, it is often desirable to make density plots. However, the hidden angular measure $S^{(l)}(\cdot)$ is supported on $\delta \mathcal{N}^{(l)}$, which is a difficult plotting domain. For example, when $d = 3$, the set $\delta \mathcal{N}^{(2)}$ is a disjoint union of six rectangles lying on three different planes as shown in Figure 4.1. Though $\delta \mathcal{N}^{(l)}$ is a $(d - 1)$-dimensional set, $d$-dimensional vectors are needed to represent $\delta \mathcal{N}^{(l)}$. So, the density plots on $\delta \mathcal{N}^{(l)}$ also requires an additional dimension. In the two dimensional case, the problem is resolved by taking a transformation of points from $\delta \mathcal{N}^{(1)} = \{ x \in \mathbb{E} : x^{(1)} = 1 \}$ to $[0, 1]$ and looking at the density of the induced probability measure of the transformed points [47, pages 316-321]. We seek similar appropriate transformations in higher dimensional cases. We devise a transformation of points from $\delta \mathcal{N}^{(l)}$ to the $(d - 1)$-dimensional simplex $\Delta_{d-1} = \{ x \in [0, 1]^{d-1} : \sum_{i=1}^{d-1} x^i \leq 1 \}$ (see Section 4.4.1). The probability measure $\tilde{S}^{(l)}(\cdot)$ on the transformed points induced by $S^{(l)}(\cdot)$ is called the transformed (hidden) angular measure. Since the set $\Delta_{d-1}$ is represented by $(d - 1)$-dimensional vectors, the problem of incorporating an additional dimension in the density plots vanishes.

For characterizations of hidden regular variation [39] it is useful to know if $\nu^{(l)}(\{ x \in \mathbb{E}^{(l)} : ||x|| > 1 \})$ is finite or not, where $||x||$ is any norm of $x$. Such knowledge is also useful for estimating probabilities of some risk sets. For example, if $\nu^{(l)}(\{ x \in \mathbb{E}^{(l)} : ||x|| > 1 \})$ is finite, then so is $\nu^{(l)}(\{ x \in \mathbb{E}^{(l)} : a_1 x^1 + a_2 x^2 + \cdots + a_d x^d > y \}, a_i > 0, i = 1, 2, \cdots, d, y > 0$. We show that this issue can be resolved by checking a moment condition.
Figure 4.1: The set $\delta \mathbb{N}^{(2)}$ (shaded region), when $d = 3$.

### 4.1.1 Outline

Section 4.1.2 explains notation. In Section 4.2, we review the definitions of regular variation on $\mathbb{E}$ and hidden regular variation on $\mathbb{E}^{(2)}$, and extend the concept to the sub-cones $\mathbb{E}^{(l)} = [0, \infty)^d \setminus \bigcup_{1 \leq j_1 < j_2 < \ldots < j_d \leq l} \{ x^{j_1} = 0, \ldots, x^{j_d-1} = 0 \}$, $3 \leq l \leq d$. Section 4.3 discusses exploratory detection techniques for hidden regular variation on $\mathbb{E}^{(l)}$ and estimation of the limit measure $\nu^{(l)}(\cdot)$. We consider in Section 4.4 a transformation that allows us to visualize the hidden angular measure $S^{(l)}(\cdot)$ through another probability measure $\tilde{S}^{(l)}(\cdot)$ on the $(d - 1)$-dimensional simplex. In Section 4.5, we discuss conditions for $\nu^{(l)}(\{ x \in \mathbb{E}^{(l)} : \| x \| > 1 \})$ being finite or not. Section 4.6 gives examples of risk sets where hidden regular variation helps in obtaining finer estimates of their probabilities. Our methodologies are applied to two examples in Section 4.7. We conclude with some remarks and outline open issues in Section 4.8.
4.1.2 Notation

Vectors and cones

For denoting a vector and its components, we use:

\[ \mathbf{x} = (x^1, x^2, \cdots, x^d), \quad x^i = i\text{-th component of } \mathbf{x}, \quad i = 1, 2, \cdots, d. \]

The vectors of all zeros, all ones and all infinities are denoted by \( \mathbf{0} = (0, 0, \cdots, 0) \), \( \mathbf{1} = (1, 1, \cdots, 1) \) and \( \infty = (\infty, \infty, \cdots, \infty) \) respectively. Operations on and between vectors are understood componentwise. In particular, for non-negative vectors \( \mathbf{x} \) and \( \mathbf{\beta} = (\beta^1, \beta^2, \cdots, \beta^d) \), write \( \mathbf{x}^{\mathbf{\beta}} = ((x^1)^{\beta^1}, (x^2)^{\beta^2}, \cdots, (x^d)^{\beta^d}) \). We denote the norm of \( \mathbf{x} \) by \( \| \mathbf{x} \| \). Unless specified, this could be taken as any norm. For the \( i \)-th largest component of \( \mathbf{x} \), we use:

\[ x^{(i)} = i\text{-th largest component of } \mathbf{x}, \quad i = 1, 2, \cdots, d, \quad \text{that is} \quad x^{(1)} \geq x^{(2)} \geq \cdots \geq x^{(d)}. \]

So, the superscripts denote components of a vector and the ordered component is denoted by a parenthesis in the superscript.

Sometimes, we have to sort the \( i \)-th largest components of the vectors \( \mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_n \) in non-increasing order. We first obtain the vector \( \{Z_1^{(i)}, Z_2^{(i)}, \cdots, Z_n^{(i)}\} \) by taking the \( i \)-th largest component for each \( \mathbf{Z}_j \) and then sort these to get

\[ Z_{(1)}^{(i)} \geq Z_{(2)}^{(i)} \geq \cdots \geq Z_{(n)}^{(i)}. \]

We use the parentheses in the subscript to avoid double parentheses on the superscript.

The cones we consider are

\[ \mathbb{E} = \mathbb{E}^{(1)} = [0, \infty)^d \setminus \{ \mathbf{0} \} = [0, \infty)^d \setminus \{ x^1 = 0, \cdots, x^d = 0 \}. \]
and for \(2 \leq l \leq d\),
\[
\mathbb{E}^{(l)} = [0, \infty)^d \setminus \bigcup_{1 \leq j_1 < j_2 < \cdots < j_{d-l+1} \leq d} \{ x^{j_1} = 0, \ldots, x^{j_{d-l+1}} = 0 \}
\]
\[
= [0, \infty)^d \setminus \{ x \in [0, \infty)^d : x^{(l)} > 0 \}.
\]

For \(2 \leq l \leq d\), \(\mathbb{E}^{(l)}\) is the set of points in \(\mathbb{E}\) such that at least \(l\) components are positive. Sometimes \(\mathbb{E}^{(2)}\) is expressed as \(\mathbb{E}^{(2)} = \mathbb{E} \setminus \bigcup_{i=1}^{d} \mathbb{L}_i\), where \(\mathbb{L}_i := \{ t \mathbf{e}_i, t > 0 \}\) is the \(i\)-th axis and \(\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)\), where 1 is in the \(i\)-th position, \(i = 1, 2, \ldots, d\). For \(x \in \mathbb{E}\), we use \([0, x]^c\) to mean
\[
[0, x]^c = \mathbb{E} \setminus [0, x] = \{ y \in \mathbb{E} : \forall_{i=1}^{d} y^i / x^i > 1 \}.
\]

### Regular variation and vague convergence.

We express vague convergence [47, page 173] of Radon measures as \(\Rightarrow\) and weak convergence of probability measures [7, page 14] as \(\Rightarrow\). Denote the set of non-negative Radon measures on a space \(\mathbb{F}\) as \(M_+(\mathbb{F})\) and the set of all non-negative continuous functions with compact support from \(\mathbb{F}\) to \(\mathbb{R^+}\) as \(C^+_{K}(\mathbb{F})\). The notation \(RV_{\rho}\) means the family of one dimensional regularly varying functions with exponent of variation \(\rho\) ([47, page 24], [8, 18]). For any measure \(m(\cdot)\) and a real-valued function \(f(\cdot)\), denote the integral \(\int f(x)m(dx)\) by \(m(f)\).

For defining regular variation of distributions of random vectors on \(\mathbb{E} = \mathbb{E}^{(1)}\) as in (1.3.1), we use the scaling function \(b(t) = b^{(1)}(t)\) and get the limit measure \(\nu(\cdot) = \nu^{(1)}(\cdot)\). Similarly, for defining regular variation of distributions of random vectors on \(\mathbb{E}^{(l)}, 2 \leq l \leq d\), we use the scaling function \(b^{(l)}(t)\) and get the limit measure \(\nu^{(l)}(\cdot)\). For each \(1 \leq l \leq d\), define the set \(\mathcal{N}^{(l)}\) by \(\mathcal{N}^{(l)} = \{ x \in \mathbb{E}^{(l)} : x^{(l)} \geq 1 \} \).
Since $\mathbf{N}^{(l)}$ is compact in $\mathbb{E}^{(l)}$ and $\cup_{t>0}t\mathbf{N}^{(l)} = \mathbb{E}^{(l)}$, there always exists a suitable choice of the scaling function $b^{(l)}(t)$ which makes $\nu^{(l)}(\mathbf{N}^{(l)}) = 1$. We assume this from now on.

For each $2 \leq l \leq d$, if we have hidden regular variation on $\mathbb{E}^{(l)}$, the limit measure $\nu^{(l)}(\cdot)$ can be expressed in a convenient coordinate system as a product of a Pareto measure $\nu_{\alpha^{(l)}}(dr) = \alpha^{(l)}r^{-\alpha^{(l)}-1}dr$, $r > 0$ and a probability measure $S^{(l)}(\cdot)$ on the compact set $\mathbf{N}^{(l)} = \{x \in \mathbb{E}^{(l)} : x^{(l)} = 1\}$. The measure $S^{(l)}(\cdot)$ is called the hidden angular or hidden spectral measure on $\mathbb{E}^{(l)}$. Whenever the limit measure $\nu^{(l)}(\cdot)$ satisfies $\nu^{(l)}(\{x \in \mathbb{E}^{(l)} : x^{(l)} \geq 1, x^{(1)} = \infty\}) = 0$, we view $S^{(l)}(\cdot)$ through its transformed version denoted $\tilde{S}^{(l)}(\cdot)$, which is a probability measure on the $(d-1)$-dimensional simplex $\Delta_{d-1} = \{x \in [0, 1]^{d-1} : \sum_{i=1}^{d-1} x^i \leq 1\}$.

Anti-ranks.

Suppose $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n$ are random vectors in $[0, \infty)^d$. For $j = 1, 2, \ldots, d$, $i = 1, 2, \ldots, n$, define the anti-rank

$$ r^j_i = \sum_{j=1}^{n} 1_{|Z^j_i \geq Z^j_i|} $$

for $Z^j_i$ to be the number of $j$-th components greater than or equal to $Z^j_i$. For $2 \leq l \leq d$, define

$$ m^{(l)}_i = \text{the } l\text{-th largest component of } (\frac{1}{r^j_i}, j = 1, 2, \ldots, d) $$

and then order them as

$$ m^{(l)}_{(1)} \geq m^{(l)}_{(2)} \geq \cdots \geq m^{(l)}_{(n)}.$$
4.2 Hidden regular variation

We give more details about regular variation on $\mathbb{E}$ and HRV on $\mathbb{E}^{(2)}$ and then extend the definitions to hidden regular variation on sub-cones of $\mathbb{E}^{(2)}$. We illustrate with some examples.

4.2.1 Hidden regular variation on $\mathbb{E}^{(2)}$

Consider regular variation on $\mathbb{E}$ and hidden regular variation on $\mathbb{E}^{(2)}$.

The standard case

The distribution of $Z = (Z^1, Z^2, \ldots, Z^d)$ is regularly varying on $\mathbb{E} := [0, \infty]^d \setminus \{0\}$ with limit measure $\nu(\cdot)$ if there exist a function $b(t) \uparrow \infty$ as $t \to \infty$ and a non-negative non-degenerate Radon measure $\nu(\cdot) \neq 0$ such that

$$
tP \left( \frac{Z}{b(t)} \in \cdot \right) \xrightarrow{\nu} \nu(\cdot) \quad \text{in } M_+(\mathbb{E}). \tag{4.2.1}
$$

The limit measure $\nu(\cdot)$ must have all non-zero marginals. Then, there exists $\alpha > 0$ such that $b(\cdot) \in RV_{1/\alpha}$ and $\nu(\cdot)$ satisfies the scaling property

$$
\nu(c\cdot) = c^{-\alpha} \nu(\cdot), \quad c > 0. \tag{4.2.2}
$$

Call the limit relation (4.2.1) the standard case which requires the same scaling function $b(t)$ for all the components of $Z$ in (4.2.1) and ensures that $\nu(\cdot)$ has all non-zero marginals.

HRV allows for another regular variation on a sub-cone such as $\mathbb{E}^{(2)}$. The distribution of $Z$ has hidden regular variation on $\mathbb{E}^{(2)}$ if in addition to (4.2.1)
There exist a non-decreasing function \( b^{(2)}(t) \uparrow \infty \) such that \( b(t)/b^{(2)}(t) \to \infty \) and a non-negative Radon measure \( \nu^{(2)}(\cdot) \neq 0 \) on \( \mathbb{B}^{(2)} \) such that

\[
t P \left[ \frac{Z_j}{b^{(2)}(t)} \in \cdot \right] \xrightarrow{\nu} \nu^{(2)}(\cdot) \quad \text{in} \ M_+ \left( \mathbf{E}^{(2)} \right); \tag{4.2.3}
\]

see [47, page 324]. It follows from (4.2.3) that there exists \( a^{(2)} \geq a \) such that \( b^{(2)}(\cdot) \in RV_{1/a^{(2)}} \) and \( \nu^{(2)}(\cdot) \) satisfies the scaling property

\[
\nu^{(2)}(c \cdot) = c^{-a^{(2)}} \nu^{(2)}(\cdot), \quad c > 0. \tag{4.2.4}
\]

HRV implies \( \nu(\mathbb{B}^{(2)}) = 0 \), which is known as asymptotic independence [47, page 324]. We emphasize that the model of hidden regular variation on \( \mathbb{B}^{(2)} \) requires both (4.2.1) and (4.2.3) to be satisfied with \( b(t)/b^{(2)}(t) \to \infty \), and not only regular variation on \( \mathbb{B}^{(2)} \) as in (4.2.3).

**The non-standard case**

Non-standard regular variation may hold when (4.2.1) fails, but

\[
t P \left[ \left( \frac{Z_j}{a^{(t)}} \right), j = 1, 2, \cdots , d \right] \in \cdot \xrightarrow{\nu} \mu(\cdot) \quad \text{in} \ M_+ (\mathbb{B}) \tag{4.2.5}
\]

for some scaling functions \( a^1(\cdot), a^2(\cdot), \cdots , a^d(\cdot) \) satisfying \( a^j(t) \uparrow \infty \), where \( \mu(\cdot) \) is a non-negative non-zero Radon measure on \( \mathbb{B} \) [19, 52]. We assume that marginal convergences satisfy

\[
t P \left[ \frac{Z_j}{a^j(t)} \in \cdot \right] \xrightarrow{\nu} \nu^{(j)}(\cdot) \quad \text{in} \ M_+ ((0, \infty]), \tag{4.2.6}
\]

where \( \nu^{(j)}((x, \infty]) = x^{-\beta^j}, \beta^j > 0, x > 0 \). Relation (4.2.5) is equivalent to the standard convergence

\[
t P \left[ \left( \frac{a^{(t)}(Z_j)}{t} \right), j = 1, 2, \cdots , d \right] \in \cdot \xrightarrow{\nu} \nu(\cdot) \quad \text{in} \ M_+ (\mathbb{B}), \tag{4.2.7}
\]
where \( \nu(\cdot) \) satisfies the scaling property \( \nu(c\cdot) = c^{-1}\nu(\cdot), \ c > 0, \) ([45, page 277], [18, 27]). The limit measures \( \nu(\cdot) \) and \( \mu(\cdot) \) are related:

\[
\mu([0, x]^c) = \nu([0, x^d]^c), \quad x \in \mathbb{E}.
\]  

(4.2.8)

In this non-standard case, the distribution of \( Z \) has hidden regular variation on \( \mathbb{E}^{(2)} \) if, in addition to (4.2.7), there exist a non-decreasing function \( b^{(2)}(t) \uparrow \infty \), such that \( t/b^{(2)}(t) \to \infty \), and a non-negative non-zero Radon measure \( \nu^{(2)}(\cdot) \) on \( \mathbb{E}^{(2)} \) satisfying

\[
tP\left[ \frac{\left( a^{(2)}(Z^j), j = 1, 2, \ldots, d \right)}{b^{(2)}(t)} \in \cdot \right] \to \nu^{(2)}(\cdot) \quad \text{in} \ M_+(\mathbb{E}^{(2)}).
\]  

(4.2.9)

Then, there exists \( a^{(2)} \geq 1 \) such that \( b^{(2)}(\cdot) \in RV_{1/(a^{(2)})} \) and \( \nu^{(2)}(\cdot) \) satisfies the scaling property (4.2.4).

Note that (4.2.7) standardizes (4.2.5) with scaling function \( b(t) = t \), and the definition of hidden regular variation on \( \mathbb{E}^{(2)} \) in (4.2.9), is the most natural substitute for (4.2.3). This reduces the non-standard case to the standard one. Of course, we have to deal with the unknown nature of the scaling functions \( a^j(\cdot), \ j = 1, 2, \ldots, d. \)

### 4.2.2 Hidden regular variation beyond \( \mathbb{E}^{(2)} \)

For dimension \( d > 2 \), it is possible to refine HRV on \( \mathbb{E}^{(2)} \) by defining hidden regular variation on sub-cones of \( \mathbb{E}^{(2)}. \)
A reason for seeking HRV on $E^{(2)}$ is that in the presence of asymptotic independence when the limit measure $\nu(\cdot)$ puts zero mass on $E^{(2)}$, regular variation on $\mathbb{B}$ may fail to provide non-zero estimates of the probabilities of remote critical sets such as failure regions (reliability), overflow regions (hydrology), and out-of-compliance regions (environmental protection) [47, page 322]. Beyond $E^{(2)}$, if the limit measure $\nu^{(2)}(\cdot)$ in (4.2.3) puts zero mass on $E^{(3)}$ we would seek to refine HRV on $E^{(2)}$.

Consider the following thought experiment. Suppose $Z = (Z_1, Z_2, \ldots, Z_d)$ represents concentrations of a pollutant at $d$ locations and that $Z$ has a regularly varying distribution on $\mathbb{B}$ with asymptotic independence. Assume we found HRV on $E^{(2)}$ and the limit measure $\nu^{(2)}(\cdot)$ in this case satisfies $\nu^{(2)}(E^{(3)}) = 0$, so HRV on $E^{(2)}$, estimates $P(Z^{j_1} > x_1, Z^{j_2} > x_2, \ldots, Z^{j_l} > x_l)$ to be 0 for $3 \leq l \leq d$ and $1 \leq j_1 < j_2 < \cdots < j_l \leq d$. This resulting estimate seems crude and we seek a remedy by looking for finer structure of on the sub-cones $E^{(3)} \supset \cdots \supset E^{(d)}$ in a sequential manner.

Another context for HRV on $E^{(3)}$ is as a refinement of regular variation on $\mathbb{B}$ when asymptotic independence is absent. Suppose in the above thought experiment, $Z$ has a regularly varying distribution on $\mathbb{B}$ with limit measure $\nu(\cdot)$ such that $\nu(E^{(2)}) > 0$, but $\nu(E^{(3)}) = 0$. Asymptotic independence is absent, but $P(Z^{j_1} > x_1, Z^{j_2} > x_2, \ldots, Z^{j_l} > x_l)$ is estimated to be 0 for all $3 \leq l \leq d$ and $1 \leq j_1 < j_2 < \cdots < j_l \leq d$. This suggests seeking HRV on the sub-cones $E^{(3)} \supset \cdots \supset E^{(d)}$.

Examples in Section 4.2.3 show each modeling situation we considered in
the above thought experiments can happen.

We seek regular variation on the cones \( E \supset E(2) \supset E(3) \supset \cdots \supset E(d) \) in a sequential manner. If for some \( 1 \leq j \leq d \), regular variation is present on \( E(j) \), as in (1.3.1) and the limit measure \( \nu^{(j)}(\cdot) \) puts non-zero mass on \( E(l) \), \( j < l \leq d \), that is \( \nu^{(j)}(E(l)) > 0 \), then there is no need to seek HRV on any of the cones \( E(j+1) \supset \cdots \supset E(d) \). Recall the conventions that we replace \( \nu(\cdot), \alpha, E \) and \( b(t) \) by \( \nu^{(1)}(\cdot), \alpha^{(1)}, E^{(1)} \) and \( b^{(1)}(t) \) respectively.

Of course, there are other ways to nest sub-regions of \( E \) and seek regular variation but our sequential search for regular variation on the cones \( E(l); l = 2, \ldots, d \) is one structured approach to the problem of refined estimates.

**Formal definition of HRV on \( E(l) \)**

The definition proceeds sequentially and begins with the standard case. Assume that \( Z \) satisfies regular variation on \( E(1) \) as in (4.2.1) and that we have regular variation on a sub-cone \( E(j) \) with scaling function \( b^{(j)}(t) \in RV_{1/\alpha^{(j)}} \) and limiting Radon measure \( \nu^{(j)}(\cdot) \neq 0 \). For \( j < l \leq d \), further assume that \( \nu^{(j)}(E(l-1)) > 0 \) and \( \nu^{(j)}(E(l)) = 0 \). The cone \( E(l) \) could be \( E(1) \). The distribution of \( Z \) has hidden regular variation on \( E(l) \), if in addition to regular variation on \( E(j) \), there is a non-decreasing function \( b^{(l)}(t) \uparrow \infty \) such that \( b^{(j)}(t)/b^{(l)}(t) \to \infty \), and a non-negative Radon measure \( \nu^{(l)}(\cdot) \neq 0 \) on \( E(l) \) such that

\[
\mathbb{P}\left[ \frac{Z}{b^{(l)}(t)} \in \cdot \right] \Rightarrow \nu^{(l)}(\cdot) \quad \text{in } M_{+}^{(l)}.
\]  

From (4.2.10), there exists \( \alpha^{(l)} \geq \alpha^{(j)} \) such that \( b^{(l)}(\cdot) \in RV_{1/\alpha^{(l)}} \) and \( \nu^{(l)}(\cdot) \) has the scaling property

\[
\nu^{(l)}(c\cdot) = c^{-\alpha^{(l)}}\nu^{(l)}(\cdot), \quad c > 0.
\]  

119
For vague convergence on \( E^{(0)} \), it is important to identify the relatively compact sets of \( E^{(0)} \). From Proposition 6.1 of [47, page 171], the relatively compact sets of \( E^{(0)} \) are subsets of sets the form \( \{ x \in E^{(1)} : x^{j_1} > w_1, x^{j_2} > w_2, \ldots, x^{j_l} > w_l \} \) for some \( 1 \leq j_1 < j_2 < \cdots < j_l \leq d \) and for some \( w_1, w_2, \ldots, w_l > 0 \). So, for all \( \delta > 0 \), \( \{ x \in [0, \infty]^d : x^{(l)} > \delta \} \) is relatively compact in \( E^{(0)} \) and

\[
\nu^{(j)}(E^{(0)}) = 0
\]

is a necessary condition for HRV on \( E^{(0)} \).

For defining HRV in the non-standard case, assume (4.2.7) holds on \( E^{(1)} \) and the rest of the definition is the same with \( Z \) and \( b^{(1)}(t) \) replaced by \( (a^{1^-}(Z^1), a^{2^-}(Z^2), \ldots, a^{d^-}(Z^d)) \) and \( t \) respectively.

**Remark 4.2.1.** A few important remarks about hidden regular variation:

(i) The definition of hidden regular variation leading to (4.2.10) is consistent with the definition of hidden regular variation on \( E^{(2)} \).

(ii) The definition of regular variation on \( E^{(1)} \) as in (4.2.1) or (4.2.7) requires that the limit measure \( \nu^{(1)}(\cdot) \) has non-zero marginals. When defining regular variation on \( E^{(l)} \), \( 2 \leq l \leq d \), as in (4.2.10), we do not demand such a condition. For instance, \( Z = (Z^1, Z^2, Z^3) \) being regularly varying on \( E^{(2)} \) does not imply that \( (Z^1, Z^2) \) is regularly varying on \( (0, \infty)^2 \). See Example 4.2.4.

(iii) Non-standard regular variation allows each component \( Z^j \) of the random vector \( Z \) to be scaled by a possibly different scaling function \( a^j(t) \) as in (4.2.5). An alternative approach to defining regular variation on \( E^{(l)} \), \( 2 \leq l \leq d \), would allow each component \( Z^j \) of the random vector \( Z \) to be scaled
by a possibly different scaling function \( b^{(l)j}(t) \) and this would produce a more general model of HRV than the one we defined. However, we do not have a method of estimating the scaling functions \( b^{(l)j}(n/k) \); see estimation of \( b^{(l)}(n/k) \) using (4.3.6) and estimation of \( a^{j}(b^{(2)}(n/k)) \) in the non-standard case using (4.6.9).

### 4.2.3 Examples

We give examples to exhibit subtleties. Example 4.2.3 shows that non-existence of HRV on \( \mathbb{E}^{(2)} \) does not preclude HRV on \( \mathbb{E}^{(3)} \). Example 4.2.3 also shows that asymptotic independence is not a necessary condition for the presence of HRV on \( \mathbb{E}^{(3)} \). In Examples 4.2.4 and 4.2.5, we learn that HRV on \( \mathbb{E}^{(2)} \) does not imply HRV on \( \mathbb{E}^{(3)} \). In Example 4.2.5, HRV on \( \mathbb{E}^{(3)} \) fails because \( \nu^{(2)}(\mathbb{E}^{(3)}) > 0 \), but a different reason for failure holds in Example 4.2.4. In contrast, Example 4.2.2 demonstrates that HRV could be present on each of the sub-cones \( \mathbb{E}^{(l)} \), \( 2 \leq l \leq d \). Also, Example 4.2.5 shows that asymptotic independence, unlike independence, does not imply \( \nu^{(2)}(\mathbb{E}^{(3)}) = 0 \).

**Example 4.2.2.** An extension of Example 5.1 of [39]: Suppose \( Z^1, Z^2, \ldots, Z^d \) are i.i.d. Pareto(1). Then, regular variation of \( Z = (Z^1, Z^2, \ldots, Z^d) \) is present on \( \mathbb{E} \) with \( \alpha = 1 \) and HRV is present on each of the sub-cones \( \mathbb{E}^{(l)} \) with \( \alpha^{(l)} = l \), for \( 2 \leq l \leq d \).

**Example 4.2.3.** Suppose \( X \) and \( Y \) are i.i.d. Pareto(1) and \( Z = (X, 2X, Y) \), so

\[
\tau P \left[ \frac{Z}{2t} \in \cdot \right] \overset{\nu(\cdot)}{\rightarrow} \nu(\cdot) \quad \text{in} \; M_+(\mathbb{E}),
\]

and \( \nu(\cdot) \) has all non-zero marginals. However, \( Z \) does not possess asymptotic independence since \( Z^1 \) and \( Z^2 \) are not asymptotically independent [45, page 296,
Proposition 5.27] and thus HRV cannot be present on $\mathbb{E}^{(2)}$ [47, page 325, Property 9.1]. However,

$$
\nu(\mathbb{E}^{(3)}) = \lim_{w \to 0} \nu(\{ x : x^1 \wedge x^2 \wedge x^3 > w \}) = \lim_{w \to 0} \lim_{t \to \infty} tP \left[ X > tw, 2X > tw, Y > tw \right]
$$

$$
= \lim_{w \to 0} \lim_{t \to \infty} tP \left[ X > tw, Y > tw \right] = \lim_{w \to 0} t(tw)^{-1}(tw)^{-1} = 0.
$$

This suggests seeking HRV on $\mathbb{E}^{(3)}$ and indeed this holds with $b^{(3)}(t) = \sqrt{t}$ since for $w_1, w_2, w_3 > 0$,

$$
\lim_{t \to \infty} tP \left[ X > \sqrt{tw_1}, 2X > \sqrt{tw_2}, Y > \sqrt{tw_3} \right] = \lim_{t \to \infty} tP \left[ X > \sqrt{t} \left( w_1 \vee \frac{w_2}{2} \right), Y > \sqrt{tw_3} \right]
$$

$$
= \lim_{t \to \infty} \left[ \sqrt{t} \left( w_1 \vee \frac{w_2}{2} \right) \right]^{-1} (\sqrt{tw_3})^{-1} = \frac{1}{\left( w_1 \vee \frac{w_2}{2} \right) w_3}.
$$

So, for this example,

(i) Regular variation holds on $\mathbb{E}^{(1)}$ and $\mathbb{E}^{(2)}$ (since $\nu(\mathbb{E}^{(2)}) \neq 0$), HRV holds on $\mathbb{E}^{(3)}$, $\nu(\mathbb{E}^{(1)}) = \nu(\mathbb{E}^{(2)}) = \infty$, $\nu(\mathbb{E}^{(3)}) = 0$.

(ii) Asymptotic independence is absent but HRV on $\mathbb{E}^{(3)}$ is present.

**Example 4.2.4.** Example 5.2 from [39]: Let $X_1, X_2, X_3$ be i.i.d. Pareto(1) random variables. Also, let $B_1, B_2$ be i.i.d. Bernoulli random variables independent of $(X_1, X_2, X_3)$ with $P[B_i = 1] = P[B_i = 0] = 1/2$, $i = 1, 2$. Define

$$
Z = (B_2 X_1, (1 - B_2) X_2, (1 - B_1) X_3).
$$

From [39], HRV exists on the cone $\mathbb{E}^{(2)}$ with $\alpha^{(2)} = 2$ and $\nu^{(2)}(\cdot)$ concentrates on $[x^1 > 0, x^3 > 0] \cup [x^2 > 0, x^3 > 0]$. Also, $\nu^{(2)}(\{ x : x^1 > 0, x^2 > 0 \}) = 0$. Since, $\mathbb{E}^{(3)}$ is a subset of $\{ x : x^1 > 0, x^2 > 0 \}$, $\nu^{(2)}(\mathbb{E}^{(3)}) = 0$. However, HRV on $\mathbb{E}^{(3)}$ fails. The compact sets of $\mathbb{E}^{(3)}$ are contained in sets of the form $\{ x : x^1 > w^1, x^2 > w^2, x^3 > w^3 \}$ for $w^1, w^2, w^3 > 0$. Since either $Z^1$ or $Z^2$ must be zero, for any increasing function
$h(t) \uparrow \infty$, and $w^1, w^2, w^3 > 0$, we have

$$\lim_{t \to \infty} tP \left[ \frac{Z}{h(t)} \in \{ x : x^1 > w^1, x^2 > w^2, x^3 > w^3 \} \right] = 0.$$  

Hence, HRV holds on $\mathbb{B}^{(2)}$ with $b^{(2)}(t) = \sqrt{t}$, but HRV on $\mathbb{B}^{(3)}$ fails.

**Example 4.2.5.** Let $X_1, X_2$ and $X_3$ be i.i.d. Pareto(1) random variables and define $Z = ((X_1)^2 \land (X_2)^2, (X_2)^2 \land (X_3)^2, (X_1)^2 \land (X_3)^2)$. First, note that

$$tP \left[ \frac{Z}{3t} \in \cdot \right] \sim n(\cdot) \text{ in } M_+(\mathbb{E})$$

for some non-zero Radon measure $n(\cdot)$ on $\mathbb{E}$ with non-zero marginals. Also,

$$tP \left[ \frac{Z}{t^{2/3}} \in \cdot \right] \sim n^{(2)}(\cdot) \text{ in } M_+(\mathbb{E}^{(2)})$$

for a non-zero Radon measure $n^{(2)}(\cdot)$ on $\mathbb{E}^{(2)}$. So, HRV exists on $\mathbb{E}^{(2)}$ and hence, the components of $Z$ are asymptotically independent [47, page 325, Property 9.1]. For $w_1, w_2, w_3 > 0$,

$$\lim_{t \to \infty} tP \left[ \frac{Z}{t^{2/3}} \in \{ x : x^1 > w_1, x^2 > w_2, x^3 > w_3 \} \right]$$

$$= \lim_{t \to \infty} tP \left[ X_1 > t^{1/3}(w_1 \lor w_3)^{1/2}, X_2 > t^{1/3}(w_1 \lor w_2)^{1/2}, X_3 > t^{1/3}(w_2 \lor w_3)^{1/2} \right]$$

$$= \frac{1}{\sqrt{(w_1 \lor w_3) \cdot (w_1 \lor w_2) \cdot (w_2 \lor w_3)}} = n^{(2)}(\{ x : x^1 > w_1, x^2 > w_2, x^3 > w_3 \}).$$

As $\{ x : x^1 > w_1, x^2 > w_2, x^3 > w_3 \} \subset \mathbb{E}^{(3)}$, $n^{(2)}(\mathbb{E}^{(3)}) > 0$. So, for this example,

(i) HRV exists on $\mathbb{E}^{(2)}$, not on $\mathbb{E}^{(3)}$, but $Z$ is regularly varying on $\mathbb{E}^{(3)}$ in the sense of (1.3.1).

(ii) Asymptotic independence holds but $n^{(2)}(\mathbb{E}^{(3)}) > 0$. 

123
4.3 Exploratory detection and estimation techniques

Existing exploratory detection techniques for HRV on $E^{(2)}$ are valid in two dimensions. Our methods, applicable to any dimension, also allow for sequential search for HRV on $E^{(l)}$, $2 \leq l \leq d$.

We find a coordinate system in which the limit measure $\nu^{(l)}(\cdot)$ in (4.2.10) is a product of a probability measure and a Pareto measure of the form $\nu_{\alpha(l)}(\cdot)$ for some $\alpha(l) > 0$. Thus we exploit the semi-parametric nature of $\nu^{(l)}(\cdot)$ for estimation and detection.

4.3.1 Decomposition of the limit measure $\nu^{(l)}(\cdot)$

By a suitable choice of scaling function $b^{(l)}(t)$, we can and do make $\nu^{(l)}(N^{(l)}) = 1$, where $N^{(l)} = \{x \in E^{(l)} : x^{(l)} \geq 1\}$. We decompose $\nu^{(l)}(\cdot)$ into a Pareto measure $\nu_{\alpha(l)}(\cdot)$ and a probability measure $S^{(l)}(\cdot)$ on $\delta N^{(l)} = \{x \in E^{(l)} : x^{(l)} = 1\}$ called the hidden angular or hidden angular measure.

Proposition 4.3.1. The distribution of the random vector $Z$ has regular variation on $E^{(l)}$, that is it satisfies (1.3.1) with $C = E^{(l)}$ and $\chi = \nu^{(l)}$, and the condition $\nu^{(l)}(N^{(l)}) = 1$ holds iff

$$tP \left[ \left( \frac{Z^{(l)}}{b^{(l)}(t)}, \frac{Z}{Z^{(l)}} \right) \in \cdot \right] \nu_{\alpha(l)} \times S^{(l)}(\cdot) \text{ in } M_+((0, \infty] \times \delta N^{(l)}),$$

(4.3.1)

where $Z^{(l)}$ is the $l$-th largest component of $Z$. The limit measure $\nu^{(l)}(\cdot)$ and the probability measure $S^{(l)}(\cdot)$ are related by

$$\nu^{(l)}(\{x \in E^{(l)} : x^{(l)} \geq r, \frac{x}{x^{(l)}} \in \Lambda\}) = r^{-\alpha(l)} S^{(l)}(\Lambda),$$

(4.3.2)

which holds for all $r > 0$ and all Borel sets $\Lambda \subset \delta N^{(l)}$. 

124
Proof. See Section 4.9. □

**Remark 4.3.2.** Proposition 4.3.1 only assumes regular variation on $\mathbb{E}^{(l)}$, $\alpha^{(l)} > 0$, whereas hidden regular variation on $\mathbb{E}^{(l)}$ also requires (4.2.1) to hold and $b(t)/b^{(l)}(t) \to \infty$.

Also, the convergence in (4.3.1) is equivalent to

(i) $Z^{(l)}$ having regularly varying tail with index $\alpha^{(l)} > 0$ and

(ii) as $t \to \infty$,

$$P \left[ \frac{Z}{Z^{(l)}} \in \cdot \mid Z^{(l)} > t \right] \Rightarrow S^{(l)}(\cdot) \quad \text{on } \delta\mathbb{N}^{(l)}.$$

**Remark 4.3.3.** The polar coordinate transformation $x \mapsto (\|x\|, x/\|x\|)$ usually used for regular variation introduces a non-compact unit sphere \{ $x \in \mathbb{E}^{(l)} : \|x\| = 1$ \}. This defect is fixed by using $\delta\mathbb{N}^{(l)}$ instead.

Example 4.3.4 uses Proposition 4.3.1 to construct random variables having regular variation on the cone $\mathbb{E}^{(l)}$ with the limit measure $\nu^{(l)}(\cdot)$.

**Example 4.3.4.** Suppose $(R, \Theta)$ is an independent pair of random variables on $(0, \infty) \times \delta\mathbb{N}^{(l)}$ with

$$P[R > r] = r^{-\alpha^{(l)}}, \quad r > 1, \quad P[\Theta \in \cdot] = S^{(l)}(\cdot).$$

Then,

$$tP\left[ \frac{R}{t^{1/\alpha^{(l)}}} > r, \Theta \in \Lambda \right] = \frac{1}{r^{\alpha^{(l)}}} S^{(l)}(\Lambda) = r^{-\alpha^{(l)}} S^{(l)}(\Lambda).$$

By Proposition 4.3.1, the distribution of $Z = R\Theta$ is regularly varying on $\mathbb{E}^{(l)}$ and satisfies (4.2.10) with $\nu^{(l)}(\mathbb{N}^{(l)}) = 1$. This, however, does not guarantee regular variation on $\mathbb{E}$. Also, unless $\Theta$ has a support contained in $\{ \theta \in \delta\mathbb{N}^{(l)} : \theta^{(l)} < \infty \}$, the components of the random vector $Z$ might not be real-valued.
4.3.2 Detection of HRV on $\mathbb{E}^{(l)}$ and estimation of $\nu^{(l)}(\cdot)$

Is the model of hidden regular variation on $\mathbb{E}^{(l)}$ appropriate for a given data set? If so, how do we estimate the limit measure $\nu^{(l)}(\cdot)$ and tail probabilities of the form $P[Z_{i_1} > z_1, Z_{i_2} > z_2, \cdots, Z_{i_l} > z_l]$ for $1 \leq i_1 < i_2 \cdots < i_l \leq d$. We consider the standard and non-standard cases and assume $\nu^{(l)}(\mathbb{N}^{(l)}) = 1$.

The standard case

Suppose $Z_1, Z_2, \cdots, Z_n$ are i.i.d. random vectors in $[0, \infty)^d$ whose common distribution satisfies regular variation on $\mathbb{E}$ as in (4.2.1). We want to detect if HRV is present in $\mathbb{E}^{(l)}$ and this requires prior detection of regular variation on a bigger sub-cone $\mathbb{E}^{(j)} \supset \mathbb{E}^{(l)}$ with the limit measure $\nu^{(j)}(\cdot)$ having the property $\nu^{(j)}(\mathbb{E}^{(l-1)}) > 0$ and $\nu^{(j)}(\mathbb{E}^{(l)}) = 0$. Recall $\mathbb{E}^{(j)}$ could be $\mathbb{E}^{(1)}$.

Here is a method for verifying that $\nu^{(j)}(\mathbb{E}^{(l-1)}) > 0$ and $\nu^{(j)}(\mathbb{E}^{(l)}) = 0$. For each $p > j$, define a transformation $J^{(p)} : \delta \mathbb{N}^{(j)} \mapsto [0, 1]$ as $x \mapsto x^{(p)}$. If $\nu^{(j)}(\mathbb{E}^{(l-1)}) > 0$ and $\nu^{(j)}(\mathbb{E}^{(l)}) = 0$, then the probability measure $S^{(j)} \circ J^{(l-1)^{-1}}(\cdot)$ is degenerate at zero but $S^{(j)} \circ J^{(l-1)^{-1}}(\cdot)$ is not; see Remark 4.4.2. As will be discussed later, we can construct an atomic measure $\hat{S}^{(j)}(\cdot)$, which consistently estimates $S^{(j)}(\cdot)$. Using the atoms of $\hat{S}^{(j)} \circ J^{(l-1)^{-1}}(\cdot)$, we plot a kernel density estimate of the density of $S^{(j)} \circ J^{(l-1)^{-1}}(\cdot)$. If the plotted density appears to concentrate around zero, we believe that $\nu^{(j)}(\mathbb{E}^{(l-1)}) = 0$. Otherwise, we assume that $\nu^{(j)}(\mathbb{E}^{(l-1)}) > 0$. Then, using similar methods, we proceed to check whether $\nu^{(j)}(\mathbb{E}^{(l)}) = 0$.

Once convinced that $\nu^{(j)}(\mathbb{E}^{(l-1)}) > 0$ and $\nu^{(j)}(\mathbb{E}^{(l)}) = 0$, we seek HRV on $\mathbb{E}^{(l)}$. 

126
Using Proposition 4.3.1, HRV implies

$$tP \left[ Z^{(l)}/b^{(l)}(t) \in \cdot \right] \Rightarrow \nu_{a^{(l)}}(\cdot) \quad \text{in } M_+(0, \infty).$$  \hspace{1cm} (4.3.3)

So, we apply Hill, QQ and Pickands plots to the i.i.d. data \( \{Z_i^{(l)}, i = 1, 2, \ldots, n\} \) and attempt to infer that \( Z^{(l)} \) has a regularly varying distribution [47, Chapter 4].

If convinced that HRV is present, we estimate the limit measure \( \nu^{(l)}(\cdot) \). Define the set

\[
\mathbb{E}_{l,\infty} = \mathbb{E}^{(l)} \setminus \cup_{1 \leq j_1 < j_2 < \cdots < j_d} \left[ x^{j_1} = \infty, x^{j_2} = \infty, \cdots, x^{j_d} = \infty \right] = \mathbb{E}^{(l)} \setminus \left[ x^{(l)} = \infty \right]
\]

and the transformation \( Q^{(l)} : \mathbb{E}_{l,\infty} \mapsto (0, \infty) \times \delta^{N^{(l)}} \) as

\[
Q^{(l)}(x) = \left(x^{(l)}, x/x^{(l)}\right).
\]  \hspace{1cm} (4.3.4)

From (4.3.2) and the fact that \( Q^{(l)}(\cdot) \) is one-one, we get for any Borel set \( A \subset \mathbb{E}^{(l)}, \)

\[
\nu^{(l)}(A) = \nu^{(l)}(A \cap \mathbb{E}_{l,\infty}) = \nu_{a^{(l)}} \times S^{(l)}(Q^{(l)}(A \cap \mathbb{E}_{l,\infty})).
\]

So, estimating \( a^{(l)} \) and the hidden angular measure \( S^{(l)}(\cdot) \) is equivalent to estimating \( \nu^{(l)}(\cdot) \).

We estimate \( a^{(l)} \) using one dimensional methods such as the Hill, QQ or Pickands estimator applied to the i.i.d. data \( \{Z_i^{(l)}, i = 1, 2, \ldots, n\} \). An estimator of \( S^{(l)}(\cdot) \) can be constructed using standard ideas [27] as follows. Suppose \( k = k(n) \to \infty, k(n)/n \to 0, \) as \( n \to \infty \). Using Theorem 5.3(ii) of [47, page 139], we get

\[
\frac{1}{k} \sum_{i=1}^{n} \mathcal{E}\left[Z_i^{(l)}/b^{(l)}(Z_i^{(l)})\right](\cdot) \Rightarrow \nu_{a^{(l)}} \times S^{(l)}(\cdot)
\]  \hspace{1cm} (4.3.5)

on \( M_+((0, \infty) \times \delta^{N^{(l)}}) \). Choosing \((1, \infty) \times \cdot\) as the set in (4.3.5), gives an estimator of \( S^{(l)}(\cdot) \), but this estimator uses the unknown \( b^{(l)}(n/k) \), which must be replaced by a statistic.
Order the observations \( \{Z_i, i = 1, 2, \cdots, n\} \) as \( Z_{(1)} \geq Z_{(2)} \geq \cdots \geq Z_{(n)} \) which are order statistics from a sample drawn from a regularly varying distribution.

Using (4.3.3) and Theorem 4.2 of [47, page 81], we get

\[
\frac{Z_{(k)}}{b(n/k)} \overset{p}{\to} 1. \tag{4.3.6}
\]

Then (4.3.5) and (4.3.6) yield

\[
\left( \frac{1}{k} \sum_{i=1}^{n} \epsilon(Z_{i(1)}^{(0)}/b(n/k), Z_{i(2)}^{(0)}) \right) \left( \frac{Z_{(k)}}{b(n/k)} \right) \Rightarrow \left( \nu_{a(0)} \times S^{(0)}(\cdot), 1 \right) \tag{4.3.7}
\]

on \( M_{+}((0, \infty] \times \delta \mathbb{N}^{(0)}) \times (0, \infty] \). Applying the almost surely continuous map

\[(v \times S, x) \mapsto v \times S((x, \infty] \times \cdot)\]

to (4.3.7), the continuous mapping theorem [7, page 21] gives

\[
\frac{1}{k} \sum_{i=1}^{n} \epsilon(Z_{i(1)}^{(0)}/b(n/k), Z_{i(2)}^{(0)}) \left( \frac{Z_{(k)}}{b(n/k)} \right) \Rightarrow \nu_{a(0)}(1, \infty] \times \cdot \times \cdot \tag{4.3.8}
\]

\[
\frac{1}{k} \sum_{i=1}^{n} \epsilon(Z_{i(1)}^{(0)}/b(n/k), Z_{i(2)}^{(0)}) \Rightarrow \nu_{a(0)}((1, \infty) \times \cdot) S^{(0)}(\cdot) = S^{(0)}(\cdot)
\]

on \( \delta \mathbb{N}^{(0)} \). Thus, a consistent estimator for \( S^{(0)}(\cdot) \) is \( \hat{S}^{(0)}(\cdot) := \frac{\sum_{i=1}^{n} \epsilon(Z_{i(1)}^{(0)}/b(n/k), Z_{i(2)}^{(0)}) \times (1, \infty]}{\sum_{i=1}^{n} \epsilon(Z_{i(1)}^{(0)}/b(n/k))} \), \( \hat{S}^{(0)}(\cdot) \)

assuming the denominator on the right side of (4.3.9) is positive. Note that the denominator is zero if \( k = n \). However, since \( k \sum_{i=1}^{n} \epsilon(Z_{i(1)}^{(0)}/b(n/k)) \) converges to \( 1 \) in probability as \( k = k(n) \to \infty, n \to \infty, n/k \to \infty \), for large \( k \) and \( n/k \), there is a negligible probability that the denominator is zero. So, on the set where the denominator is zero, we can set \( \hat{S}^{(0)}(\cdot) \) arbitrarily as any probability measure on \( \delta \mathbb{N}^{(0)} \), for example \( \hat{S}^{(0)}(\cdot) = \epsilon_1(\cdot) \).
The non-standard case

Suppose $Z_1, Z_2, \ldots, Z_n$ are i.i.d. random vectors in $[0, \infty)^d$ such that their common distribution satisfies non-standard regular variation (4.2.7) on $\mathbb{E}$. We seek HRV on $\mathbb{E}(l)$. HRV is defined sequentially, so if HRV on $\mathbb{E}(l)$ exists,

(i) either (4.2.7) holds, and $\left(a^+(Z^i), i = 1, 2, \cdots, d\right)$ is standard regularly varying on $\mathbb{E}(l), \mathbb{E}(l)$ with limit measures $\nu(l)(\cdot), \nu(l)(\cdot)$ and scaling functions $b(l)(t), b(l)(t)$ for $1 \leq j < l \leq d$ and $\nu(l)(\mathbb{E}(l-1)) > 0$ and $\nu(l)(\mathbb{E}(l)) = 0$,

(ii) or (4.2.7) holds, (4.2.10) holds with $Z$ replaced by $\left(a^+(Z^i), i = 1, 2, \cdots, d\right)$, and $\nu(\mathbb{E}(l-1)) > 0$ and $\nu(\mathbb{E}(l)) = 0$.

In each case, (4.2.7) holds, (4.2.10) holds with $\left(a^+(Z^i), i = 1, 2, \cdots, d\right)$ replacing $Z$, and $t/b(l)(t) \to \infty$.

Recall the definitions from Section 4.1.2 of antiranks $\{r_i^l, i = 1, 2, \cdots, n, j = 1, 2, \cdots, d\}$, $l$-th largest components of $\{1/r_i^l, j = 1, 2, \cdots, d\}$ denoted $m_i(l)$ for each $i$, and order statistics of $\{m_i(l), i = 1, 2, \cdots, n\}$, denoted $\{m_p(l), p = 1, 2, \cdots, n\}$. Here is a method to detect HRV on $\mathbb{E}(l)$ in the non-standard case.

**Proposition 4.3.5.** Assume that $Z_1, Z_2, \ldots, Z_n$ are i.i.d. random vectors from a distribution on $[0, \infty)^d$ that satisfies both regular variation on $\mathbb{E}$ and HRV on $\mathbb{E}(l)$, so that (4.2.7) holds and (4.2.10) holds with $Z$ replaced by $\left(a^+(Z^j), j = 1, 2, \cdots, d\right)$.

We assume that $\nu(l)(\mathbb{N}(l)) = 1$. Then, we have on $M_+(\mathbb{E}(l))$,

$$\hat{\nu}(l)(\cdot) := \frac{1}{k} \sum_{i=1}^n \epsilon_{\left(1/r_i^l/m_{(1)}^{(l)}, 1 \leq j \leq d\right)}(\cdot) \Rightarrow \nu(l)(\cdot) \text{ on } M_+(\mathbb{E}(l)). \quad (4.3.10)$$

**Proof.** For $l = d$, the statement is the same as Proposition 2 of [27], except that instead of defining HRV on $\mathbb{E}(2)$, we have assumed HRV on $\mathbb{E}(d)$. The proof of the case $2 \leq l < d$ is similar to the case for $l = d$ and is omitted. \[\square\]
Remark 4.3.6. In the case, \( l = 2 < d \), the only improvement of Proposition 4.3.5, over Proposition 2 of [27] is that here we assume \( \nu^{(2)(\mathcal{N}^{(2)})} = 1 \) instead of assuming \( \nu^{(2)(\{x \in \mathbb{B}^{(2)} : \land_{j=1}^{d} x_j \geq 1\})} = 1 \). We claim that if HRV on \( \mathbb{B}^{(2)} \) is present, the assumption \( \nu^{(2)(\mathcal{N}^{(2)})} = 1 \) could always be achieved by a suitable choice of \( b^{(2)}(t) \), but if \( d > 2 \), this may not be true for the assumption of \( \nu^{(2)(\{x \in \mathbb{B}^{(2)} : \land_{j=1}^{d} x_j \geq 1\})} = 1 \), as claimed in [27]. See Example 4.2.2 for an illustration.

Proposition 4.3.5 gives us a consistent estimator of \( \nu^{(l)(\cdot)} \), without using the semi-parametric structure of \( \nu^{(l)(\cdot)} \) resulting from (4.2.11) and we now exploit this structure. In the non-standard case, decomposition of \( \nu^{(l)(\cdot)} \) is achieved as in Proposition 4.3.1, only the role of \( Z \) is played by \((a_j \leftarrow (Z_j), j = 1, 2, \ldots, d)\). The limit measure \( \nu^{(l)(\cdot)} \) of (4.2.10) is related to the hidden angular measure \( S^{(l)(\cdot)} \) through (4.3.2), which acts as the definition of the hidden angular measure \( S^{(l)(\cdot)} \) in the non-standard case.

Proposition 4.3.7. The following two statements are equivalent:

1. The estimator of \( \nu^{(l)(\cdot)} \) based on ranks is consistent as \( k = k(n) \to \infty \), \( k(n)/n \to 0 \), and \( n \to \infty \); that is

\[
\hat{\nu}^{(l)(\cdot)} := \frac{1}{k} \sum_{i=1}^{n} \epsilon_{(1/r_j)/m_{ij}, 1 \leq j \leq d)}(\cdot) \Rightarrow \nu^{(l)(\cdot)} \quad \text{on } M_{+}(\mathbb{B}^{(l)}). \tag{4.3.11}
\]

2. The estimator of \( \nu_{\alpha^{(l)}} \times S^{(l)(\cdot)} \) based on ranks is consistent as \( k = k(n) \to \infty \), \( k(n)/n \to 0 \), and \( n \to \infty \); that is

\[
\frac{1}{k} \sum_{i=1}^{n} \epsilon_{(m_{ij}^{(l)}, m_{ij}^{(l)})/((1/r_j)/m_{ij}^{(l)}, 1 \leq j \leq d)}(\cdot) \Rightarrow \nu_{\alpha^{(l)}} \times S^{(l)(\cdot)} \quad \text{on } M_{+}((0, \infty] \times \delta_{N^{(l)}}). \tag{4.3.12}
\]

Proof. See Section 4.9. \( \Box \)
Detection of hidden regular variation on $E^{(l)}$, for some $2 \leq l \leq d$, requires the prior conclusion that $(a^i Z_i, i = 1, 2, \cdots, d)$ is standard regularly varying on a bigger sub-cone $E^{(j)} \supset E^{(l)}$. Using the rank transform, we explore for regular variation on $E$ and then move sequentially through the cones $E \supset E^{(2)} \supset \cdots$. We also need $\nu^{(j)}$ to satisfy $\nu^{(j)}(E^{(l-1)}) > 0$ and $\nu^{(j)}(E^{(l)}) = 0$ which is verified using the hidden angular measure $S^{(l)}\cdot$. Finally, we verify regular variation on the cone $E^{(l)}$. From Proposition 4.3.5 and Proposition 4.3.7, HRV on $E^{(l)}$ implies

$$\frac{1}{k} \sum_{i=1}^{n} \epsilon_{m_i^{(j)}/m_{i\delta}} \Rightarrow \nu_{\alpha^{(j)}} \text{ on } M_+((0, \infty]). \quad (4.3.13)$$

We can use, for example, a Hill plot to determine whether (4.3.13) is true since consistency of the Hill estimator is only dependent on the consistency of the tail empirical measure and does not require the tail empirical measure to be constructed using i.i.d. data. ([53], [47, page 80]). This gives us an exploratory method for detecting hidden regular variation on $E^{(l)}$ in the non-standard case.

To estimate the limit measure $\nu^{(l)}\cdot$, it is again sufficient to estimate $\alpha^{(l)}$ and the hidden angular measure $S^{(l)}\cdot$. Estimate $\alpha^{(l)}$ using, say, the Hill estimator based on the rank-based data $\{m_i^{(l)}, i = 1, 2, \cdots, n\}$ [47, Chapter 4] and using Proposition 4.3.5 and Proposition 4.3.7, we get in $M_+(\delta N^{(l)})$ that

$$\frac{1}{k} \sum_{i=1}^{n} \epsilon_{m_i^{(l)}/m_{i\delta}, ((1/j)_i/m_i^{(l)}, 1 \leq j \leq d)}([1, \infty] \times \cdot) \Rightarrow \nu_{\alpha^{(l)}}([1, \infty])S^{(l)}\cdot = S^{(l)}\cdot \text{ or }$$

$$\hat{S}^{(l)}\cdot := \frac{\sum_{i=1}^{n} \epsilon_{m_i^{(l)}/m_{i\delta}, ((1/j)_i/m_i^{(l)}, 1 \leq j \leq d)}([1, \infty] \times \cdot)}{\sum_{i=1}^{n} \epsilon_{m_i^{(l)}/m_{i\delta}}([1, \infty])} \Rightarrow S^{(l)}\cdot. \quad (4.3.14)$$

This gives a consistent estimator of $S^{(l)}\cdot$. 

131
4.4 A different representation of the hidden angular measure

\[ S^{(l)}(\cdot) \]

As discussed in the introduction, we map points of \( \delta \mathbb{N}^{(l)} \) to the \((d - 1)\)-dimensional simplex \( \Delta_{d-1} = \{ x \in [0,1]^{d-1} : \sum_{i=1}^{d-1} x^i \leq 1 \} \). The probability measure \( \tilde{S}^{(l)}(\cdot) \) on the transformed points induced by \( S^{(l)}(\cdot) \) is called the transformed (hidden) angular measure. However, we must make the standing assumption that

\[ \nu^{(l)}(\{ x \in \mathbb{B}^{(l)} : x^{(l)} = \infty \}) = 0, \quad \text{for all } 2 \leq l \leq d, \quad (4.4.1) \]

whenever \( \nu^{(l)}(\cdot) \) exists, since otherwise the transformation is not one-one. Assumption (4.4.1) is not very strong and most examples satisfy this assumption. Nonetheless, this assumption is not always true, as illustrated by examples in Section 4.4.4. Recall the conventions that we replace \( \nu(\cdot), \alpha, S(\cdot) \) and \( \tilde{S}(\cdot) \) by \( \nu^{(1)}(\cdot), \alpha^{(1)}, S^{(1)}(\cdot) \) and \( \tilde{S}^{(1)}(\cdot) \) respectively.

4.4.1 The transformation

First note that \( \nu^{(1)}(\{ x \in \mathbb{B}^{(1)} : x^{(1)} = \infty \}) = 0 \) due to the scaling property of \( \nu^{(1)}(\cdot) \) in (4.2.2) and the compactness of \( \{ x \in \mathbb{B}^{(l)} : x^{(l)} \geq 1 \} \) in \( \mathbb{B}^{(l)} \). So we may modify (4.4.1) to include \( l = 1 \).

Define a transformation \( T^{(l)} : \delta \mathbb{N}^{(l)} \mapsto \Delta_{d-1} := \{ s \in [0,1]^{d-1} : \sum_{i=1}^{d-1} s^i \leq 1 \} \) for each \( l, 1 \leq l \leq d \), which is one-one on an appropriate subset of \( \delta \mathbb{N}^{(l)} \). The appropriate subset is

\[ D_{1}^{(l)} := \{ x \in \delta \mathbb{N}^{(l)} : x^{(1)} < \infty \}. \quad (4.4.2) \]
On $D_1^{(l)}$, define $T^{(l)}(\cdot)$ as

$$T^{(l)}(x) = \frac{(x^2, x^3, \ldots, x^d)}{\sum_{i=1}^d x^i}.$$  \hspace{1cm} (4.4.3)

To identify $T^{(l)}(D_1^{(l)})$, first we define a map $\phi^{(l)} : \Delta_{d-1} \mapsto [0, 1]$ as

$$\phi^{(l)}(s^1, s^2, \ldots, s^{d-1}) = \text{the } l\text{-th largest component of } (1 - \sum_{i=1}^{d-1} s^i, s^2, \ldots, s^{d-1}).$$  \hspace{1cm} (4.4.4)

Using this notation, we see that

$$D_2^{(l)} := T^{(l)}(D_1^{(l)}) = \{(s^1, s^2, \ldots, s^{d-1}) \in \Delta_{d-1} : \phi^{(l)}(s^1, s^2, \ldots, s^{d-1}) > 0\} \subset \Delta_{d-1}. \hspace{1cm} (4.4.5)$$

To show that $T^{(l)}(\cdot)$ is one-one on $D_1^{(l)}$, we explicitly define the inverse transformation $T^{(l)-1} : D_2^{(l)} \mapsto D_1^{(l)}$ as

$$T^{(l)-1}(s^1, s^2, \ldots, s^{d-1}) = \frac{(1 - \sum_{i=1}^{d-1} s^i, s^2, \ldots, s^{d-1})}{\phi^{(l)}(s^1, s^2, \ldots, s^{d-1})}. \hspace{1cm} (4.4.6)$$

We extend our definition of $T^{(l)}(\cdot)$ from $D_1^{(l)}$ to the entire set $\delta\mathbb{N}^{(l)}$ by setting $T^{(l)}(x) = 0$ for $x \in D_1^{(l)} \setminus D_1^{(l)c}$. We define a similar extension of $T^{(l)-1}(\cdot)$ to the whole simplex $\Delta_{d-1}$ by setting $T^{(l)-1}(s^1, s^2, \ldots, s^{d-1}) = 1$ for $(s^1, s^2, \ldots, s^{d-1}) \in D_2^{(l)c}$. Now define the probability measure $\tilde{S}^{(l)}(\cdot) = S^{(l)} \circ T^{(l)-1}(\cdot)$ on $\Delta_{d-1}$; this is called the transformed hidden angular measure on $\mathbb{E}^{(l)}$. Note that, by assumption (4.4.1),

$$S^{(l)}(D_1^{(l)c}) = \nu^{(l)}(\{x \in \mathbb{E}^{(l)} : x^{(l)} \geq 1, \frac{x}{x^{(l)}} \in D_1^{(l)c}\}) = \nu^{(l)}(\{x \in \mathbb{E}^{(l)} : x^{(l)} \geq 1, x^{(l)} = \infty\}) = 0.$$

Therefore, using (4.4.5), we get $\tilde{S}^{(l)}(D_2^{(l)}) = 1$. Since $T^{(l)}(\cdot)$ is one-one on $D_1^{(l)}$ and $S^{(l)}(D_1^{(l)}) = 1$, for any Borel set $A \subset \delta\mathbb{N}^{(l)}$, we can compute $S^{(l)}(A)$ by noting that

$$S^{(l)}(A) = S^{(l)}(A \cap D_1^{(l)}) = \tilde{S}^{(l)}(T^{(l)}(A \cap D_1^{(l)})). \hspace{1cm} (4.4.7)$$

So, studying the transformed hidden angular measure $\tilde{S}^{(l)}(\cdot)$ on the nice set $\Delta_{d-1}$ is sufficient to understand the hidden angular measure $S^{(l)}(\cdot)$. 

4.4.2 Estimation of $\tilde{S}^{(l)}(\cdot)$

In the standard case, we get from (4.3.9),

$$\hat{S}^{(l)}(\cdot) := \frac{1}{k} \sum_{i=1}^{n} \epsilon(z_i^{(l)}/z_{ij}^{(l)}, z_i/z_i^{(l)})((1, \infty] \times \cdot) \Rightarrow S^{(l)}(\cdot)$$ (4.4.8)

on $M_+(\delta N^{(l)})$. The function $T^{(l)}(\cdot)$ defined in (4.4.3) is continuous on $D_1^{(l)}$ and hence is continuous almost surely with respect to the probability measure $S^{(l)}(\cdot)$. Therefore, by the continuous mapping theorem [7, page 21],

$$\hat{S}^{(l)} \circ T^{(l)-1}(\cdot) := \frac{1}{k} \sum_{i=1}^{n} \epsilon(z_i^{(l)}/z_{ij}^{(l)}, T^{(l)}(z_i/z_i^{(l)}))((1, \infty] \times \cdot) \Rightarrow S^{(l)} \circ T^{(l)-1}(\cdot) = \tilde{S}^{(l)}(\cdot)$$ (4.4.9)

on $M_+(\Delta_{d-1})$. Conversely, (4.4.9) implies (4.4.8) by continuity of $T^{(l)-1}(\cdot)$ on $D_2^{(l)}$ and the fact $\tilde{S}^{(l)}(D_2^{(l)}) = 1$. Thus (4.4.8) and (4.4.9) are equivalent.

In the non-standard case, (4.3.14) implies that on $M_+(\delta N^{(l)})$,

$$\tilde{S}^{(l)}(\cdot) := \frac{1}{k} \sum_{i=1}^{n} \epsilon(m_i^{(l)}/m_{i\alpha}^{(l)}, (1/r_j^{(l)})/m_i^{(l)}, 1 \leq j \leq d))((1, \infty] \times \cdot) \Rightarrow S^{(l)}(\cdot).$$

By a similar argument as in the standard case, this is equivalent to the fact that on $M_+(\Delta_{d-1})$,

$$\hat{S}^{(l)} \circ T^{(l)-1}(\cdot) := \frac{1}{k} \sum_{i=1}^{n} \epsilon(m_i^{(l)}/m_{i\alpha}^{(l)}, T^{(l)}((1/r_j^{(l)})/m_i^{(l)}, 1 \leq j \leq d))((1, \infty] \times \cdot) \Rightarrow S^{(l)} \circ T^{(l)-1}(\cdot) = \tilde{S}^{(l)}(\cdot).$$

4.4.3 Supports of transformed (hidden) angular measure $\tilde{S}^{(l)}(\cdot)$

The following lemma illustrates that the supports of the transformed (hidden) angular measures are disjoint.

**Lemma 4.4.1.** Recall $D_2^{(l)}$ defined in (4.4.5). For $1 \leq j < l \leq d$,

$$\nu^{(l)}(D_2^{(l)}) = 0 \text{ iff } \tilde{S}^{(l)}(D_2^{(l)}) = 0.$$
Proof. By the scaling property (4.2.2) or (4.2.11), \( \nu^j(\{x \in \mathbb{E} : x^j = \infty\}) = 0 \), and hence, by the continuous mapping theorem,

\[
\nu^j(E^{(j)}) = \nu^j(E^{(j)} \cap \{x \in \mathbb{E} : x^j < \infty\}) = \nu_{\alpha^j} \times S^j(Q^{(j)}(E^{(j)} \cap \{x \in \mathbb{E} : x^j < \infty\}),
\]

where \( Q^{(j)}(x) = \left( x^j, \frac{x}{x^j} \right) \). Now,

\[
\nu_{\alpha^j} \times S^j(Q^{(j)}(E^{(j)} \cap \{x \in \mathbb{E} : x^j < \infty\})) = \nu_{\alpha^j} \times S^j((r, \theta) \in (0, \infty) \times \delta N^j : \theta^j > 0)\]

\[
\quad = \lim_{\lambda \to 0} \lambda^{-\alpha^j} S^j((\theta \in \delta N^j : \theta^j > 0)).
\]

Hence, \( \nu^j(E^{(j)}) = 0 \) iff \( S^j((\theta \in \delta N^j : \theta^j > 0)) = 0 \). Since \( S^j(D_1^{(j)}) = 1 \), where \( D_1^{(j)} \) is as given in (4.4.2), we get

\[
S^j((\theta \in \delta N^j : \theta^j > 0)) = S^j((\theta \in \delta N^j : \theta^j > 0) \cap D_1^{(j)})
\]

\[
\quad = \tilde{S}^j(T^j((\theta \in \delta N^j : \theta^j > 0) \cap D_1^{(j)}))
\]

\[
\quad = \tilde{S}^j((s^1, s^2, \ldots, s^{d-1}) \in \Delta_{d-1} : \phi^j(s^1, s^2, \ldots, s^{d-1}) > 0))
\]

\[
\quad = \tilde{S}^j(D_2^{(j)}).
\]

Hence, the result follows. \( \square \)

Remark 4.4.2. The fact that \( \nu^j(E^{(j)}) = 0 \) iff \( S^j((\theta \in \delta N^j : \theta^j > 0)) = 0 \), follows from the proof of Lemma 4.4.1. Notice, this result does not require the assumption (4.4.1).

If \( \nu^j(E^{(j)}) = 0 \) and HRV on \( E^{(j)} \) exists, then the support of \( \tilde{S}^j(\cdot) \) is contained in \( D_2^{(j)c} \) and the support of \( \tilde{S}^j(\cdot) \) is contained in \( D_2^{(j)} \), which are disjoint. So, if one seeks (hidden) regular variation on the nested cones \( \mathbb{E} = \mathbb{E}^{(1)} \supset \mathbb{E}^{(2)} \supset \cdots \supset \mathbb{E}^{(d)} \), if HRV is present, the transformed angular measure and the transformed hidden angular measures on \( \Delta_{d-1} \) will have disjoint supports.

For a visual illustration, fix \( d = 3 \) and suppose \( \tilde{S}^{(1)} \) is concentrated on the corner points of the triangle \( \Delta_2 \). By Lemma 4.4.1, \( \nu^{(1)}(E^{(2)}) = 0 \) and we search for
HRV on $\mathbb{R}^2$. Assume that it is indeed present and so consider $\tilde{S}^{(2)}$. As we have already noticed, the support of $\tilde{S}^{(2)}$ is contained in $D_2^{(2)}$ and hence does not put any mass on the corner points of the triangle $\Delta_2$. Therefore, $\tilde{S}^{(2)}$ and $\tilde{S}^{(1)}$ have disjoint supports. Two cases might arise from this situation. In the first case, $\tilde{S}^{(2)}$ puts positive mass in the interior of the triangle $\Delta_2$. Applying Lemma 4.4.1, we infer that $\nu^{(2)}(\mathbb{E}^{(3)}) > 0$ which rules out the possibility of HRV on $\mathbb{E}^{(3)}$. Hence, we do not consider $\tilde{S}^{(3)}$. In the second case, $\tilde{S}^{(2)}$ is concentrated on the axes of the triangle $\Delta_2$ and by Lemma 4.4.1, $\nu^{(2)}(\mathbb{E}^{(3)}) = 0$. Hence, as usual, we search for HRV on $\mathbb{E}^{(3)}$ and let us assume that it is present. Then, we consider $\tilde{S}^{(3)}$. As noted, the support of $\tilde{S}^{(3)}$ is contained in $D_2^{(3)}$ and hence it only puts mass in the interior of the triangle $\Delta_2$. Hence, in this case, all three of $\tilde{S}^{(1)}$, $\tilde{S}^{(2)}$ and $\tilde{S}^{(3)}$ have disjoint supports.

Now, consider another case, where $\tilde{S}^{(1)}$ is not concentrated on the corner points of the triangle $\Delta_2$, but is concentrated on its axes. Using Lemma 4.4.1, $\nu(\mathbb{E}^{(2)}) > 0$, but $\nu^{(1)}(\mathbb{E}^{(3)}) = 0$. So, we should not search for HRV on $\mathbb{E}^{(2)}$ and hence should not consider $\tilde{S}^{(2)}$. However, we consider presence of HRV on $\mathbb{E}^{(3)}$ and hence consider $\tilde{S}^{(3)}$. But, the support of $\tilde{S}^{(3)}$ is contained in the interior of the triangle $\Delta_2$ and hence $\tilde{S}^{(3)}$ does not put any mass on the axes. So, in this case also, we would consider only $\tilde{S}^{(3)}$ and $\tilde{S}^{(1)}$, which have disjoint supports.

In the final case, suppose $\tilde{S}^{(1)}$ puts mass in the interior of the triangle $\Delta_2$. Lemma 4.4.1 implies $\nu^{(1)}(\mathbb{E}^{(3)}) > 0$ and we should not seek HRV on any of the sub-cones $\mathbb{E}^{(2)}$ or $\mathbb{E}^{(3)}$.

In all these illustrative cases, the transformed angular measure and the transformed hidden angular measures have disjoint supports.
### 4.4.4 Lines through $\infty$

Section 4.4 made the standing assumption (4.4.1), which is not always true. In Example 4.4.3, the measure $\nu^{(2)}(\cdot)$ concentrates on the lines through $\infty$; that is, on the set $\{x \in \mathbb{E}^{(2)} : x^{(1)} = \infty\}$. Examples 4.4.4 and 4.4.5 show that for $2 \leq j < l \leq d$, $\nu^{(l)}(\{x \in \mathbb{E}^{(l)} : x^{(l)} \geq 1, x^{(1)} = \infty\}) = 0$ does not imply $\nu^{(j)}(\{x \in \mathbb{E}^{(j)} : x^{(j)} \geq 1, x^{(1)} = \infty\}) = 0$ and vice versa.

**Example 4.4.3.** Let $X$ and $Y$ be two i.i.d. Pareto(1) random variables. Let $B$ be another random variable independent of $(X, Y)$ such that $P[B = 0] = P[B = 1] = \frac{1}{2}$. Define

$$Z = (Z^1, Z^2) = B(X, X^2) + (1 - B)(Y^2, Y),$$

so that

$$tP\left[\frac{Z}{t^2} \in \cdot \right] \xrightarrow{\nu} \nu(\cdot) \quad \text{in } M_+(\mathbb{E}),$$

where for $w_1, w_2 > 0$, $\nu([w_1, \infty) \times [0, \infty)) = \frac{1}{2}w_1^{-1/2}$, $\nu([0, \infty) \times (w_2, \infty)) = \frac{1}{2}w_2^{-1/2}$ and $\nu(\mathbb{E}^{(2)}) = 0$. For $w_1, w_2 > 0$,

$$\lim_{t \to \infty} tP\left[\frac{Z}{t} \in (w_1, \infty) \times (w_2, \infty)\right] = \lim_{t \to \infty} \frac{t}{2}P[X > tw_1, X^2 > tw_2] + \lim_{t \to \infty} \frac{t}{2}P[Y^2 > tw_1, Y > tw_2]$$

$$= \lim_{t \to \infty} \frac{t}{2}P[X > tw_1] + \lim_{t \to \infty} \frac{t}{2}P[Y > tw_2] = \frac{1}{2} \left( \frac{1}{w_1} + \frac{1}{w_2} \right).$$

So HRV exists on the cone $\mathbb{E}^{(2)}$ with limit measure $\nu^{(2)}(\cdot)$ such that

$$\nu^{(2)}((w_1, \infty) \times (w_2, \infty)) = \frac{1}{2} \left( \frac{1}{w_1} + \frac{1}{w_2} \right).$$

Hence, letting $w_2 \to \infty$, we get $\nu^{(2)}((w_1, \infty) \times \{\infty\}) = \frac{1}{2w_1}$ and similarly, $\nu^{(2)}((\{\infty\} \times (w_2, \infty)) = \frac{1}{2w_2}$. So, we conclude that in this case, $\nu^{(2)}(\{x \in \mathbb{E}^{(l)} : x^{(2)} \geq 1, x^{(1)} = \infty\}) = 1$.

**Example 4.4.4.** Let $X_1, X_2, \cdots, X_5$ be five i.i.d. Pareto(1) random variables. Let $(B_1, B_2, B_3)$ be another set of random variables independent of $(X_1, X_2, \cdots, X_5)$
such that \( P[B_i = 1] = 1 - P[B_i = 0] = \frac{1}{3} \) and \( \sum_{i=1}^{3} B_i = 1 \). Now, define \( Z \) as

\[
Z = (Z^1, Z^2, Z^3) = B_1(X_1, X_1^2, 0) + B_2(X_2^2, X_2, 0) + B_3(X_3^2, X_4, X_5^2).
\]

It follows that

\[
tP \left[ \frac{Z}{25t^2/9} \in \cdot \right] \to \nu(\cdot) \quad \text{in} \quad M_+(\mathbb{B}),
\]

where for \( w_1, w_2, w_3 > 0 \), \( \nu((w_1, \infty] \times [0, \infty] \times [0, \infty]) = \frac{2}{5}w_1^{-1/2}, \nu([0, \infty] \times [w_2, \infty] \times [0, \infty]) = \frac{2}{5}w_2^{-1/2}, \nu([0, \infty] \times [0, \infty] \times (w_3, \infty)) = \frac{1}{5}w_3^{-1/2} \) and \( \nu(\mathbb{B}^{(2)}) = 0 \). Now, we look for HRV on \( \mathbb{B}^{(2)} \). Notice that

\[
tP \left[ \frac{Z}{5t/3} \in \cdot \right] \to \nu^{(2)}(\cdot) \quad \text{in} \quad M_+(\mathbb{B}^{(2)}),
\]

where for \( w_1, w_2, w_3 > 0 \), \( \nu^{(2)}((0, \infty] \times (w_2, \infty] \times (w_3, \infty)) = \frac{1}{5}(w_2w_3)^{-1/2}, \nu^{(2)}((w_1, \infty] \times [0, \infty] \times (w_3, \infty)) = \frac{1}{5}(w_3w_1)^{-1/2} \) and \( \nu^{(2)}(\mathbb{B}^{(3)}) = 0 \). Hence, letting \( w_2 \to \infty \), we get

\[
\nu^{(2)}((w_1, \infty] \times [0, \infty]) = \frac{1}{5w_1},
\]

and so \( \nu^{(2)}(\{x \in \mathbb{B}^{(2)} : x^{(2)} \geq 1, x^{(1)} = \infty\}) > 0 \). We now seek HRV on the cone \( \mathbb{B}^{(3)} \).

For \( w_1, w_2, w_3 > 0 \),

\[
\lim_{t \to \infty} tP \left[ \frac{Z}{(t/3)^{2/3}} \in \cdot \right] = \lim_{t \to \infty} \frac{t}{3}P \left[ X_2^2 > (t/3)^{2/3}w_1, X_4^2 > (t/3)^{2/3}w_2, X_5^2 > (t/3)^{2/3}w_3 \right] = (w_1w_2w_3)^{-1/2}.
\]

So, HRV exists on the cone \( \mathbb{B}^{(3)} \) with limit measure \( \nu^{(3)}(\cdot) \) so that for \( w_1, w_2, w_3 > 0 \),

\[
\nu^{(3)}((w_1, \infty] \times (w_2, \infty] \times (w_3, \infty)) = (w_1w_2w_3)^{-1/2}.
\]

Hence, for this example, \( \nu^{(3)}(\{x \in \mathbb{B}^{(3)} : x^{(3)} \geq 1, x^{(1)} = \infty\}) = 0 \) and thus for

\[ 2 \leq j < l \leq d, \quad \nu^{(j)}(\{x \in \mathbb{B}^{(j)} : x^{(j)} \geq 1, x^{(1)} = \infty\}) = 0 \]

does not imply \( \nu^{(j)}(\{x \in \mathbb{B}^{(j)} : x^{(j)} \geq 1, x^{(1)} = \infty\}) = 0 \).
Example 4.4.5. Let $X_1, X_2, \ldots, X_5$ be five i.i.d. Pareto(1) random variables. Let $(B_1, B_2, B_3)$ be another set of random variables independent of $(X_1, X_2, \ldots, X_5)$ such that $P[B_i = 1] = 1 - P[B_i = 0] = \frac{1}{3}$ and $\sum_{i=1}^{3} B_i = 1$. Now, define $Z$ as

$$Z = (Z^1, Z^2, Z^3) = B_1(X_1, X_1^3, X_1^{5/4}) + B_2(X_2^3, X_2, X_2^{5/4}) + B_3(X_3^3, X_4^3, X_5^3).$$

It follows that

$$tP\left[\frac{Z}{125t^{3/27}} \in \cdot \right] \xrightarrow{\nu} \nu(\cdot) \quad \text{in} \ M_+(\mathbb{E}),$$

where for all $w_1, w_2, w_3 > 0, \nu((w_1, \infty] \times [0, \infty] \times [0, \infty]) = \frac{2}{3}w_1^{-1/3}, \nu([0, \infty] \times (w_2, \infty] \times [0, \infty]) = \frac{2}{3}w_2^{-1/3}, \nu((0, \infty] \times [0, \infty] \times (w_3, \infty)) = \frac{1}{3}w_3^{-1/3}$ and $\nu(\mathbb{E}(2)) = 0$. Now, when we seek HRV on $\mathbb{E}(2)$, we get

$$tP\left[\frac{Z}{t^{3/2}} \in \cdot \right] \xrightarrow{\nu^{(2)}} \nu^{(2)}(\cdot) \quad \text{in} \ M_+(\mathbb{E}(2)),$$

where for $w_1, w_2, w_3 > 0, \nu^{(2)}((w_1, \infty] \times (w_2, \infty] \times [0, \infty]) = \frac{1}{3}(w_1w_2)^{-1/3}, \nu^{(2)}([0, \infty] \times (w_2, \infty] \times (w_3, \infty)) = \frac{1}{3}(w_2w_3)^{-1/3}, \nu^{(2)}((0, \infty] \times [0, \infty] \times (w_3, \infty)) = \frac{1}{3}(w_3w_1)^{-1/3}$ and $\nu^{(2)}(\mathbb{E}(3)) = 0$. Notice, $\nu^{(2)}(\{x \in \mathbb{E}(2) : x^{(2)} \geq 1, x^{(1)} = \infty\}) = 0$. Now, we look for HRV on the cone $\mathbb{E}(3)$. For $w_1, w_2, w_3 > 0$,

$$\lim_{t \to \infty} tP\left[\frac{Z}{t} \in (w_1, \infty] \times (w_2, \infty] \times (w_3, \infty)\right]$$

$$= \lim_{t \to \infty} tP\left[X_1 > tw_1, X_1^3 > tw_2, X_1^{5/4} > tw_3\right]$$

$$+ \lim_{t \to \infty} tP\left[X_2^3 > tw_1, X_2 > tw_2, X_2^{5/4} > tw_3\right]$$

$$+ \lim_{t \to \infty} tP\left[X_3^3 > tw_1, X_4^3 > tw_2, X_5^3 > tw_3\right]$$

$$= \lim_{t \to \infty} \frac{t}{3} P[X_1 > tw_1] + \lim_{t \to \infty} \frac{t}{3} P[X_2 > tw_2]$$

$$+ \lim_{t \to \infty} \frac{t}{3} P[X_3 > (tw_1)^{1/3}, X_4 > (tw_2)^{1/3}, X_5 > (tw_3)^{1/3}]$$

$$= \frac{1}{3} \left(w_1^{-1} + w_2^{-1} + (w_1w_2w_3)^{-1/3}\right).$$

So, HRV exists on the cone $\mathbb{E}(3)$ with limit measure $\nu^{(3)}(\cdot)$ such that

$$\nu^{(3)}((w_1, \infty] \times (w_2, \infty] \times (w_3, \infty)) = \frac{1}{3} \left(w_1^{-1} + w_2^{-1} + (w_1w_2w_3)^{-1/3}\right).$$
Following Example 4.4.3, \( \nu^{(3)}(\{x \in \mathbb{E}^{(3)} : x^{(3)} \geq 1, x^{(1)} = \infty\}) = 2/3 \) so that for \( 2 \leq j < l \leq d \), \( \nu^{(j)}(\{x \in \mathbb{E}^{(j)} : x^{(j)} \geq 1, x^{(1)} = \infty\}) = 0 \) does not imply \( \nu^{(l)}(\{x \in \mathbb{E}^{(l)} : x^{(l)} \geq 1, x^{(1)} = \infty\}) = 0 \).

### 4.5 Deciding finiteness of \( \nu^{(l)}(\{x \in \mathbb{E}^{(l)} : ||x|| > 1\}) \)

For characterizations of HRV [39], it is useful to characterize when \( \nu^{(l)}(\{x \in \mathbb{E}^{(l)} : ||x|| > 1\}) \) is finite, where \( ||x|| \) is any norm of \( x \). Such characterizations are also useful for estimating risk set probabilities. For example, the limit measure \( \nu^{(l)}(\cdot) \) puts finite mass on a risk set of the form \( \{x \in \mathbb{E}^{(l)} : a_{1}x^{1} + a_{2}x^{2} + \cdots + a_{d}x^{d} > y\}, a_{i} > 0, i = 1, 2, \cdots, d, y > 0 \), iff \( \nu^{(l)}(\{x \in \mathbb{E}^{(l)} : ||x|| > 1\}) \) is finite. The HRV theory is not useful for estimation of risk set probability if the limit measure puts infinite mass on that risk region.

The following section resolves this issue using a moment condition. Subsequently we show that for \( 2 \leq j < l \leq d \), neither \( \nu^{(j)}(\{x \in \mathbb{E}^{(j)} : ||x|| > 1\}) \) being finite implies \( \nu^{(j)}(\{x \in \mathbb{E}^{(j)} : ||x|| > 1\}) \) is finite, nor the reverse is true.

#### 4.5.1 A moment condition

The following theorem gives a necessary and sufficient condition for the finiteness of \( \nu^{(l)}(\{x \in \mathbb{E}^{(l)} : ||x|| > 1\}) \). For \( d = 2 \), the condition of Theorem 4.5.1 is given in Proposition 5.1 of [39].

**Theorem 4.5.1.** For each \( l, 2 \leq l \leq d \), the limit measure \( \nu^{(l)}(\cdot) \) puts finite mass on
the set \( \{ x \in \mathbb{E}(l) : \|x\| > 1 \} \), that is \( \nu^{(l)}(\{ x \in \mathbb{E}(l) : \|x\| > 1 \}) \) is finite iff

\[
\int_{\delta\mathbb{N}_0} \|\theta\|^{\alpha(l)} S^{(l)}(d\theta) < \infty. \tag{4.5.1}
\]

**Proof.** We have,

\[
\nu^{(l)}(\{ x \in \mathbb{E}(l) : \|x\| > 1 \}) = \nu^{(l)}(\{ x \in \mathbb{E}(l) : x^{(l)}||x^{(l)}|| > 1 \})
\]

\[
= \nu_{\alpha(l)} \times S^{(l)}(\{ (r, \theta) \in (0, \infty) \times \delta\mathbb{N}(l) : r||\theta|| > 1 \})
\]

\[
= \int_{\delta\mathbb{N}_0} \nu_{\alpha(l)}(\{ r \in (0, \infty) : r > 1/||\theta|| \}) S^{(l)}(d\theta) = \int_{\delta\mathbb{N}_0} \|\theta\|^{\alpha(l)} S^{(l)}(d\theta).
\]

Hence, the result follows. \( \square \)

The following corollaries translate the condition of Theorem 4.5.1 to the transformed hidden angular measure \( \tilde{S}^{(l)}(\cdot) \).

**Corollary 4.5.2.** If \( \nu^{(l)}(\{ x \in \mathbb{E}(l) : x^{(l)} \geq 1, x^{(1)} = \infty \}) > 0 \), then \( \nu^{(l)}(\{ x \in \mathbb{E}(l) : \|x\| > 1 \}) \) is infinite.

**Proof.** Observe, if we denote the largest component of \( \theta \) as \( \theta^{(1)} \), we get

\[
S^{(l)}(\{ \theta \in \delta\mathbb{N}(l) : \theta^{(1)} = \infty \}) = \nu^{(l)}(\{ x \in \mathbb{E}(l) : x^{(l)} \geq 1, x^{(1)} = \infty \}) > 0.
\]

Hence, the result follows from Theorem 4.5.1. \( \square \)

**Corollary 4.5.3.** Suppose \( \nu^{(l)}(\{ x \in \mathbb{E}(l) : x^{(l)} \geq 1, x^{(1)} = \infty \}) = 0 \). Then, \( \nu^{(l)}(\{ x \in \mathbb{E}(l) : \|x\| > 1 \}) \) is finite iff

\[
\int_{D_2^{(l)}} \left( \|(1 - \sum_{i=1}^{d-1} \phi^{(l)}(s^i, s^{(1)}, s^{(2)}), \ldots, s^{(d-1)})) \|^{\alpha(l)} \right) \tilde{S}^{(l)}(ds) < \infty, \tag{4.5.2}
\]

where \( \phi^{(l)} \) and \( D_2^{(l)} \) are defined in (4.4.4) and (4.4.5) respectively.
Proof. The condition \( S^{(l)}(D_1^{(l)}) = \nu^{(l)}(\{x \in \mathbb{E}^{(l)} : x^{(l)} \geq 1, x^{(1)} = \infty\}) = 0 \), where \( D_1^{(l)} \) is defined in (4.4.2), allows us to apply the change of variable formula to (4.5.1) using the almost surely one-one transformation \( T^{(l)} \) as in (4.4.3). Now, the result follows from Theorem 4.5.1. \( \square \)

Choosing the \( L_1 \)-norm in (4.5.2), we get the simple condition:

\[
\nu^{(l)}(\{x \in \mathbb{E}^{(l)} : ||x|| > 1\}) < \infty \text{ iff } \int_{D_2^{(l)}} d^{(l)}(s^1, s^2, \ldots, s^{d-1})^{-\alpha^{(l)}} \tilde{S}^{(l)}(ds) < \infty.
\]

### 4.5.2 A particular construction

We defined HRV on a series of sub-cones \( \mathbb{B} \supset \mathbb{B}^{(2)} \supset \mathbb{B}^{(3)} \supset \cdots \supset \mathbb{B}^{(d)} \), and discussed the finiteness condition in Theorem 4.5.1 for each of the limit measures \( \nu^{(l)}(\cdot) \), \( 2 \leq l \leq d \). A natural question is if for some \( 2 \leq j < l \leq d \), HRV exists on both the cones \( \mathbb{B}^{(j)} \) and \( \mathbb{B}^{(l)} \), does finiteness of \( \nu^{(j)}(\{x \in \mathbb{B}^{(j)} : ||x|| > 1\}) \) imply finiteness of \( \nu^{(l)}(\{x \in \mathbb{B}^{(l)} : ||x|| > 1\}) \) or vice versa? We construct an example to show that there are no such implications.

**Example 4.5.4.** Suppose \( X_i, i = 1, 2, \cdots, d \) are i.i.d. Pareto(1). Also, assume \( R_i, i = 2, 3, \cdots, d \) are mutually independent random variables with \( R_i \) having distribution \( \text{Pareto}\left(\frac{i(i+1)}{2i+1}\right) \). Now, for each \( 2 \leq l \leq d \), define a set of mutually independent random variables \( s_i, i = 2, 3, \cdots, d \), such that \( s_i \) has a distribution \( \tilde{S}^{(i)} \) on \( \{x \in D_2^{(i)} : x^1 = x^{i+1} = \cdots = x^{d-1} = 0\} \), where \( D_2^{(i)} \) is defined in (4.4.5). Also, assume that \( (X_i, i = 1, 2, \cdots, d) \), \( (R_i, i = 2, 3, \cdots, d) \) and \( (s_i, i = 2, 3, \cdots, d) \) are independent of each other. Note that even though we have restricted the supports of the probability measures \( \tilde{S}^{(i)} \), we still have the flexibility to choose them in a
way so that (4.5.2) is satisfied or not, depending on whether we want to make
$v(\{x \in \mathbb{E}^{(l)} : \|x\| > 1\})$ finite or infinite.

Now, let $(B_1, B_2, \cdots, B_d)$ be another set of random variables independent of
all the previous random variables such that $P[B_i = 1] = 1 - P[B_i = 0] = \frac{1}{d}$ and
$\sum_{i=1}^{d} B_i = 1$. Recall the definition of the transformation $T^{(l)-1}$ from (4.4.6), which
maps points from $D_2^{(l)}$ to $\delta N^{(l)} = \{x \in \mathbb{E}^{(l)} : x^{(l)} = 1\}$. Note that, the range of $T^{(l)-1}$ is
$D_1^{(l)}$, where $D_1^{(l)}$ is defined in (4.4.2). Now, define the random vector $Z$ as

$$Z = (Z^1, Z^2, \cdots, Z^d)$$

$$= B_1(X_1, X_2, \cdots, X_d) + B_2R_2T^{(2)-1}(s_2) + B_3R_3T^{(3)-1}(s_3) + \cdots + B_dR_dT^{(d)-1}(s_d).$$

Since the range of $T^{(l)-1}$ is $D_1^{(l)}$, all the components of $T^{(l)-1}(s_l)$ are finite, $2 \leq l \leq d,$
and hence all the components of $Z$ are $[0, \infty)$-valued. Also,

$$tP[Z/t \in \cdot] \xrightarrow{\nu(\cdot)} \nu(\cdot) \text{ in } M_*(\mathbb{E}),$$

where $\nu([0, \infty] \times \cdots \times [0, \infty] \times (u, \infty] \times [0, \infty] \times \cdots \times [0, \infty]) = (d \cdot u)^{-1}$, where $(u, \infty]$ is in the $i$-th position and this holds for all $1 \leq i \leq d$. Also, $\nu(\mathbb{E}^{(2)}) = 0$. Notice,
for each $2 \leq l \leq d$, the parameter of the distribution of $R_l$ is chosen in such a
way that HRV of $(X_1, X_2, \cdots, X_d)$ on $\mathbb{E}^{(l)}$ or regular variation of $R_pT^{(p)-1}(s_p)$ on $\mathbb{E}^{(l)}$,
$l < p \leq d$, does not affect the HRV of $Z$ on $\mathbb{E}^{(l)}$. Also, by choosing the support
of $\tilde{S}^{(p)}$, $2 \leq p \leq d$, to be concentrated on $\{x \in D_2^{(p)} : x^{p} = x^{p+1} = \cdots = x^{d-1} = 0\}$
we have ensured that $R_pT^{(p)-1}(s_p)$, $2 \leq p < l$, would not have any HRV on the
cone $\mathbb{E}^{(l)}$. So, the only part of $Z$ contributing in HRV on $\mathbb{E}^{(l)}$ is $R_lT^{(l)-1}(s_l)$, and
therefore, for $2 \leq l \leq d$ and $x > 0$,

$$\lim_{t \to \infty} tP \left[ \frac{Z^{(l)}}{(t/d)^{(2l+1)/l(l+1)}} > x, \frac{Z}{Z^{(l)}} \in \cdot \right] = \lim_{t \to \infty} \int_{(t/d)^{(2l+1)/l(l+1)}}^\infty \frac{R_l}{(t/d)^{(2l+1)/l(l+1)}} > x, T^{(l)-1}(s_l) \in \cdot \right]$$

$$= \lim_{t \to \infty} \int_{(t/d)^{(2l+1)/l(l+1)}}^\infty \frac{R_l}{(t/d)^{(2l+1)/l(l+1)}} > x \right] P\left[ T^{(l)-1}(s_l) \in \cdot \right].$$

143
Therefore, for the cones \( E \), we could construct a random variable which has regular variation on each of \( E \), could be done independently for each \( E \), flexibility to choose (4.4.2). So, we get the transformed hidden angular measure \( \tilde{\nu} \) earlier \( \tilde{\nu} \), could independently choose to make \( \nu \). Hence, following Proposition 4.3.1, for \( 2 \leq l \leq d \), \( Z \) has regular variation on the cone \( E \) with scaling function \( b^{(l)}(t) = (t/d)^{(2l+1)/(l+1)} \), \( \alpha^{(l)} = l/(l+1) \) and hidden angular measure \( S^{(l)}(\cdot) = P[T^{(l)}^{-1}(s_i) \in \cdot] \). Also, notice \( 1/\alpha^{(l)} = (1 + l/(l+1)) \) is a decreasing function in \( l \), which indeed confirms that for \( 2 \leq j < l \leq d \), \( b^{(j)}(t)/b^{(l)}(t) \rightarrow \infty \), which is a required condition for HRV on \( E \). So, \( Z \) has HRV on the each of the cones \( E \) with the limit measure \( \nu^{(l)}(\cdot) \), \( 2 \leq l \leq d \). Now, we look for the transformed hidden angular measure \( \tilde{S}^{(l)}(\cdot) \) for the limit measure \( \nu^{(l)}(\cdot) \) and show that it indeed coincides with \( \tilde{S}^{(l)}(\cdot) \).

Since the hidden angular measure \( S^{(l)}(\cdot) \) has been defined through the function \( T^{(l)}^{-1} \) which has range \( D_1^{(l)} \), we have \( S^{(l)}(D_1^{(l)}) = 1 \), where \( D_1^{(l)} \) is defined in (4.4.2). So, we get the transformed hidden angular measure \( \tilde{S}^{(l)}(\cdot) \) as \( \tilde{S}^{(l)}(\cdot) = P[s_i \in \cdot] \). So, this hidden transformed angular measure \( \tilde{S}^{(l)}(\cdot) \) matches with our earlier \( \tilde{S}^{(l)}(\cdot) \). Following the comments made before about \( \tilde{S}^{(l)}(\cdot) \), we have the flexibility to choose \( \tilde{S}^{(l)}(\cdot) \) in such a way that (4.5.2) is satisfied or not, and this could be done independently for each \( 2 \leq l \leq d \). So, this example shows that we could construct a random variable which has regular variation on each of the cones \( E \) with limit measure \( \nu^{(l)}(\cdot) \), \( 2 \leq l \leq d \), and for each \( 2 \leq l \leq d \), we could independently choose to make \( \nu^{(l)}(x \in E^{(l)} : ||x|| > 1) \) finite or infinite. Therefore, for \( 2 \leq j < l \leq d \), neither \( \nu^{(j)}(x \in E^{(j)} : ||x|| > 1) \) is finite implies \( \nu^{(l)}(x \in E^{(l)} : ||x|| > 1) \) is finite, nor the reverse is true.
4.6 Computation of probabilities of risk sets

In this section, we consider two risk regions and illustrate how HRV helps obtain more accurate estimates of probabilities of risk sets.

4.6.1 At least one risk is large.

One scenario has $Z = (Z^1, Z^2, \ldots, Z^d)$ representing risks such as pollutant concentrations at $d$ sites [27]. A critical risk level, such as pollutant concentration $t^i$ ($i = 1, 2, \cdots, d$) at the $i$-th site, is set by a government agency. Exceeding $t^i$ for some $i$ results in a fine and the event non-compliance is represented by $\cup_{i=1}^d [Z^i > t^i]$. The probability of non-compliance is,

$$P[\text{non-compliance}] = P[\cup_{i=1}^d [Z^i > t^i]] = \sum_i P[Z^i > t^i] - \sum_{1 \leq i_1 < i_2 \leq d} P[Z^{i_1} > t^{i_1}, Z^{i_2} > t^{i_2}] + \cdots + (-1)^{(l-1)} \sum_{1 \leq i_1 < i_2 < \cdots < i_l \leq d} P[Z^{i_1} > t^{i_1}, Z^{i_2} > t^{i_2}, \cdots, Z^{i_l} > t^{i_l}]$$

$$+ \cdots + (-1)^{(d-1)} P[Z^1 > t^1, Z^2 > t^2, \cdots, Z^d > t^d].$$

Suppose $Z, Z_1, Z_2, \ldots, Z_n$ are i.i.d. random vectors whose common distribution, for simplicity, is assumed standard regularly varying on $E = E^{(1)}$ with scaling function $b(t) = b^{(1)}(t)$ as in (4.2.1). Assume HRV holds on each of the cones $E^{(l)}$ with scaling function $b^{(l)}(t)$ as in (4.2.10), $2 \leq l \leq d$. Since asymptotic independence is present, relying only on regular variation on $E$ means all the interaction terms in the inclusion-exclusion formula are estimated to be 0 but HRV improves on this.

Estimating $P[Z^i > t^i], 1 \leq i \leq d$, is a standard procedure, perhaps using peaks
over threshold and maximum likelihood; see [18, page 141], [11]. For $2 \leq j \leq d$, $1 \leq i_1 < i_2 < \cdots i_j \leq d$, large $k$ and large $n/k$, the probability $P[Z^{i_1} > t^{i_1}, \cdots, Z^{i_j} > t^{i_j}]$ is approximated using HRV on $\mathbb{E}^{(j)}$ by

$$
P[Z^{i_1} > t^{i_1}, \cdots, Z^{i_j} > t^{i_j}] = P \left[ \frac{Z^{i_1}}{b^{(j)}(n/k)} > \frac{t^{i_1}}{b^{(j)}(n/k)}, \cdots, \frac{Z^{i_j}}{b^{(j)}(n/k)} > \frac{t^{i_j}}{b^{(j)}(n/k)} \right] 
= \approx \frac{k}{n} v^{(j)} \left( \left\{ x \in \mathbb{E}^{(j)} : \frac{x^{i_1}}{b^{(j)}(n/k)} > \frac{t^{i_1}}{b^{(j)}(n/k)}, \cdots, \frac{x^{i_j}}{b^{(j)}(n/k)} > \frac{t^{i_j}}{b^{(j)}(n/k)} \right\} \right).$$

(4.6.1)

We need to estimate $v^{(j)}(\cdot)$ and $b^{(j)}(n/k)$. Notice that, for $w^1, \cdots, w^j > 0$,

$$
v^{(j)} \left( \left\{ x \in \mathbb{E}^{(j)} : \frac{x^{i_1}}{w^{i_1}}, \cdots, \frac{x^{i_j}}{w^{i_j}} \right\} \right) 
= v_{\hat{\alpha}^{(j)}} \times S^{(j)} \left( \left\{ (r, \theta) \in (0, \infty) \times \delta \mathbb{X}^{(j)} : r \hat{\theta}^{i_1} > w^{i_1}, \cdots, r \hat{\theta}^{i_j} > w^{i_j} \right\} \right) 
= \int_{\delta \mathbb{X}^{(j)}} \left( v^{(j)} \prod_{p=1}^{j} \frac{w^{i_p}}{\hat{\theta}^{i_p}} \right)^{-\hat{\alpha}^{(j)}} S^{(j)}(d\theta).$$

(4.6.2)

Using (4.3.6), we get

$$
Z^{(j)}(n/k)/b^{(j)}(n/k) \xrightarrow{P} 1,
$$

(4.6.3)

and thus we use $Z^{(j)}(n/k)$ as an estimator of $b^{(j)}(n/k)$. From (4.6.1), (4.6.2) and (4.6.3), we approximate $P[Z^{i_1} > t^{i_1}, \cdots, Z^{i_j} > t^{i_j}]$ as

$$
P[Z^{i_1} > t^{i_1}, \cdots, Z^{i_j} > t^{i_j}] \approx \frac{k}{n} \int_{\delta \mathbb{X}^{(j)}} \left( \prod_{p=1}^{j} \frac{\hat{\theta}^{i_p}}{Z^{(j)}(n/k)} \right)^{-\hat{\alpha}^{(j)}} \hat{S}^{(j)}(d\theta),
$$

where $\hat{\alpha}^{(j)}$ and $\hat{S}^{(j)}(\cdot)$ are the consistent estimates of $\alpha^{(j)}$ and $S^{(j)}(\cdot)$ obtained in Section 4.3.2.

### 4.6.2 Linear combination of risks.

A second kind of risk set used in hydrology [10, 17] is of the form $\{x \in \mathbb{E} : \gamma_1x^1 + \gamma_2x^2 + \cdots + \gamma_dx^d > y\}$ for $\gamma_i > 0$, $i = 1, 2, \cdots, d$ and $y > 0$. Here the risks
could be wind speed and wave height and a linear combination represents dike exceedance. Assume for simplicity $d = 2$ and note

$$P[\gamma_1 Z^1 + \gamma_2 Z^2 > y] = P[\gamma_1 Z^1 > y] + P[\gamma_2 Z^2 > y] - P[\gamma_1 Z^1 > y, \gamma_2 Z^2 > y]$$

$$+ P[\gamma_1 Z^1 + \gamma_2 Z^2 > y, \gamma_1 Z^1 \leq y, \gamma_2 Z^2 \leq y].$$

(4.6.4)

Suppose $Z, Z_1, Z_2, \cdots, Z_n$ are i.i.d. vectors whose common distribution has non-standard regular variation on $E = E^{(1)}$ as in (4.2.7) and HRV on $E^{(2)}$ with scaling function $b^{(2)}(t)$ as in (4.2.9). Asymptotic independence holds and thus regular variation on $E$ estimates the last two terms on the right hand side of (4.6.4) as zero. This is crude and HRV should improve the risk estimate.

As in the previous scenario, estimating $P[\gamma_i Z^i > y], i = 1, 2$, using (4.2.6) is standard and we proceed to estimate $P[\gamma_1 Z^1 > y, \gamma_2 Z^2 > y]$. From Section 2.3 in [27], we have (4.2.9) equivalent to

$$tP\left[\left(\frac{Z^j}{\alpha^j(b^{(2)}(t))}, j = 1, 2\right) \in \cdot \right] \overset{y}{\rightarrow} \nu^{(2)}(\cdot) \text{ in } M_+(E^{(2)}),$$

(4.6.5)

where $\nu^{(2)}(\cdot)$ and $\nu(\cdot)$ are related by

$$\nu^{(2)}((x, \infty)) = \nu^{(2)}((x^\beta, \infty)), \ x \in E^{(2)},$$

(4.6.6)

where $\beta = (\beta^1, \beta^2)$ and $\beta^j, j = 1, 2$, is the marginal index of regular variation defined in (4.2.6). Using (4.6.5) and (4.6.6), we approximate $P[\gamma_1 Z^1 > y, \gamma_2 Z^2 > y]$ as

$$P[\gamma_1 Z^1 > y, \gamma_2 Z^2 > y]$$

(4.6.7)

$$= P \left[ \frac{Z^1}{\alpha^1(b^{(2)}(n/k))} > \frac{y}{\gamma_1 a^1(b^{(2)}(n/k))}, \frac{Z^2}{\alpha^2(b^{(2)}(n/k))} > \frac{y}{\gamma_2 a^2(b^{(2)}(n/k))} \right]$$

$$\approx \frac{k}{n} \nu^{(2)}\left( \left\{ x \in E^{(2)} : x^1 > \frac{y}{\gamma_1 a^1(b^{(2)}(n/k))}, x^2 > \frac{y}{\gamma_2 a^2(b^{(2)}(n/k))} \right\} \right)$$

$$= \frac{k}{n} \nu^{(2)}\left( \left\{ x \in E^{(2)} : x^1 > \left( \frac{y}{\gamma_1 a^1(b^{(2)}(n/k))} \right)^{\beta^1}, x^2 > \left( \frac{y}{\gamma_2 a^2(b^{(2)}(n/k))} \right)^{\beta^2} \right\} \right).$$
We require estimates of $\nu^2(\cdot), \beta^i$ and $a^i(b^2(n/k)), i = 1, 2$. There are standard methods for estimating one dimensional indices $\beta^i, i = 1, 2$, based on (4.2.6) ([47, Chapter 4], [11, 18]) which yield consistent estimators $\hat{\beta}^i, i = 1, 2$. For $\nu^2(\cdot)$, observe,

$$\nu^2\left(\left\{ x \in \mathcal{B}^2 : x^1 > w^1, x^2 > w^2 \right\}\right) = \nu_{a(2)} \times S^{(2)}\left(\left\{ (r, \theta) \in (0, \infty] \times \delta \mathcal{N}^{(2)} : r \theta^1 > w^1, r \theta^2 > w^2 \right\}\right) = \int_{\delta \mathcal{N}^{(2)}} \left( \frac{w^1}{\theta^1} \right)^{-\alpha(2)} S^{(2)}(d\theta), \quad (w^1, w^2 > 0). \quad (4.6.8)$$

Also, from Section 4.3 in [27], we get

$$Z^j_{(1/m^2_{(k)})}/a^j\left( b^2(n/k) \right) \overset{p}{\to} 1, \quad (4.6.9)$$

where $Z^j_{(1/m^2_{(k)})}$ is the $[1/m^2_{(k)}]$-th largest order statistic of the $j$-th components of $Z_i, i = 1, 2, \cdots, n$. So, we use $Z^j_{(1/m^2_{(k)})}$ as an estimator of $a^j\left( b^2(n/k) \right), j = 1, 2$.

Finally, using (4.6.7), (4.6.8) and (4.6.9), we approximate $P[\gamma_1Z^1 > y, \gamma_2Z^2 > y]$ as

$$P[\gamma_1Z^1 > y, \gamma_2Z^2 > y] \approx \frac{k}{n} \int_{\delta \mathcal{N}^{(2)}} \left( \frac{\gamma^2}{\theta^2} \right)^{\gamma^2_{a(2)}} S^{(2)}(d\theta), \quad (4.6.10)$$

where $\gamma^{(2)}_a$ and $\gamma^{(2)}(\cdot)$ are consistent estimates of $a^{(2)}$ and $S^{(2)}(\cdot)$ obtained in Section 4.3.2.

Estimation of the fourth term of the right side of (4.6.4) requires care. First, observe

$$P[\gamma_1Z^1 + \gamma_2Z^2 > y, \gamma_1Z^1 \leq y, \gamma_2Z^2 \leq y]$$

$$= P\left[ \gamma_1a^1(b^2(n/k)) \frac{Z^1}{a^1(b^2(n/k))} + \gamma_2a^2(b^2(n/k)) \frac{Z^2}{a^2(b^2(n/k))} > y, \right.$$  

$$\left. \gamma_1a^1(b^2(n/k)) \frac{Z^1}{a^1(b^2(n/k))} \leq y, \gamma_2a^2(b^2(n/k)) \frac{Z^2}{a^2(b^2(n/k))} \leq y \right]$$
\[
\approx \frac{k}{n} \tilde{v}^{(2)}(x \in \mathbb{B} : \gamma_1 a_1(b^{(2)}(n/k)) x^1 + \gamma_2 a_2(b^{(2)}(n/k)) x^2 > y,
\gamma_1 a_1(b^{(2)}(n/k)) x^1 \leq y, \gamma_2 a_2(b^{(2)}(n/k)) x^2 \leq y))
\approx \frac{k}{n} \tilde{v}^{(2)}(x \in \mathbb{B} : \gamma_1 Z^{(1)}_{(\nu_{(\bullet)}^{(\gamma_{(2)})})} x^1 + \gamma_2 Z^{(2)}_{(\nu_{(\bullet)}^{(\gamma_{(2)})})} x^2 > y, \gamma_1 Z^{(1)}_{(\nu_{(\bullet)}^{(\gamma_{(2)})})} x^1 \vee \gamma_2 Z^{(2)}_{(\nu_{(\bullet)}^{(\gamma_{(2)})})} x^2 \leq y)).
\]

(4.6.11)

In the last approximation in (4.6.11), \(a^{j}(b^{(2)}(n/k))\) is replaced by \(Z^{j}_{(\nu_{(\bullet)}^{(\gamma_{(2)})})}\), \(j = 1, 2\), using (4.6.9).

For \(\phi_1, \phi_2 > 0\), the set \(\{x \in \mathbb{B} : \phi_1 x^1 + \phi_2 x^2 > y, \phi_1 x^1 \leq y, \phi_2 x^2 \leq y\} \subset \mathbb{B}^{(2)}\) is not a compact subset of \(\mathbb{B}^{(2)}\), so, \(\tilde{v}^{(2)}(x \in \mathbb{B} : \phi_1 x^1 + \phi_2 x^2 > y, \phi_1 x^1 \leq y, \phi_2 x^2 \leq y))\) could be infinite in which case it is not clear how HRV can refine the estimate of \(P[\gamma_1 Z^{(1)} + \gamma_2 Z^{(2)} > y, \gamma_1 Z^{(1)} \leq y, \gamma_2 Z^{(2)} \leq y]\). So, we must check finiteness of the quantity on the right side of (4.6.11).

Set \(\phi_j = \gamma_j Z^{j}_{(\nu_{(\bullet)}^{(\gamma_{(2)})})}, j = 1, 2\) and define \(A := \{x \in \mathbb{B} : \phi_1 x^1 + \phi_2 x^2 > y, \phi_1 x^1 \leq y, \phi_2 x^2 \leq y\}\). Using (4.6.6) and following similar methods as in (4.6.2), we get

\[
\tilde{v}^{(2)}(x \in \mathbb{B} : \phi_1 x^1 + \phi_2 x^2 > y, \phi_1 x^1 \leq y, \phi_2 x^2 \leq y)) = \int_A \beta_1^{(2)}(x^1)(\beta_2^{(2)} - 1)(x^2)(\beta_2^{(2)} - 1) S^{(2)}(d\gamma) = \int_{\mathbb{B}^{(2)}} \beta_1^{(2)}(\theta^1)(\beta_2^{(2)} - 1)(\theta^2)(\beta_2^{(2)} - 1) S^{(2)}(d\theta) = \int_{\mathbb{B}^{(2)}} \frac{\beta_1^{(2)}(\theta^1)(\beta_2^{(2)} - 1)(\theta^2)(\beta_2^{(2)} - 1)}{\beta_1^{(2)} + \beta_2^{(2)} - \alpha^{(2)} - 2(\theta^1)(\beta_2^{(2)} - 1) - \left(\frac{y}{\phi_1^{(2)} + \phi_2^{(2)}}\right)^{(\beta_1^{(2)} + \beta_2^{(2)} - \alpha^{(2)} - 2)}} S^{(2)}(d\theta).
\]

(4.6.12)

Finiteness of the quantity on the right hand side of (4.6.11) is equivalent to the finiteness of the quantity on the right hand side of (4.6.12) which is difficult to verify; see [27]. This problem is inherent in estimation for this type of risk region.
We proceed assuming the finiteness of \( \tilde{\nu}^{(2)}([x \in E : \phi_1 x^1 + \phi_2 x^2 > y, \phi_1 x^1 \leq y, \phi_2 x^2 \leq y]) \). From (4.6.11) and (4.6.12), we get for large \( k \) and \( n/k \), the estimate,

\[
P[y_1 Z^1 + y_2 Z^2 > y, \gamma_1 Z^1 \leq y, \gamma_2 Z^2 \leq y] \approx \frac{k}{n} \int_{\mathbb{R}^2} \frac{\hat{\beta}_1 \hat{\beta}_2}{\hat{\beta}_1 + \hat{\beta}_2 - \hat{\alpha}^{(2)} - 2} \left( \hat{\beta}_1^{\hat{\beta}_1 - 1} \hat{\beta}_2^{\hat{\beta}_2 - 1} \right)
\times \left[ \left( \frac{y}{\phi_1 \theta^1 + \phi_2 \theta^2} \right)^{\hat{\beta}_1 + \hat{\beta}_2 - \hat{\alpha}^{(2)} - 2} - \left( \frac{y}{\phi_1 \theta^1 \vee \phi_2 \theta^2} \right)^{\hat{\beta}_1 + \hat{\beta}_2 - \hat{\alpha}^{(2)} - 2} \right] \hat{S}^{(2)}(d\theta),
\]

where \( \hat{\alpha}^{(2)} \) and \( \hat{S}^{(2)}(\cdot) \) are consistent estimates of \( \alpha^{(2)} \) and \( S^{(2)}(\cdot) \) obtained in Section 4.3.2.

### 4.7 Computational examples

This section considers the performance of the estimation procedure described in Section 4.6 on two data sets, one simulated and one consisting of Internet measurements. We also compare performance with Heffernan and Resnick [27].

#### 4.7.1 Simulated data

We simulated i.i.d. samples \( \{ (X_i, Y_i), i = 1, 2, \cdots, n = 5000 \} \), where \( X_1 \sim \text{Pareto}(1) \), \( Y_1 \sim \text{Pareto}(2) \) and \( X_1 \) and \( Y_1 \) are independent. Therefore, using (4.2.9) we get

\[
\nu^{(2)}((x, y), \infty) = \frac{1}{xy}, \quad (x, y > 0), \quad (4.7.1)
\]

and \( \alpha^{(2)} = 2 \) and \( \nu^{(2)}([x \in \mathbb{E}^{(2)} : x^{(2)} \geq 1, x^{(1)} = \infty]) = 0 \). Using (4.3.2) and (4.4.7), we obtain the transformed hidden angular measure is

\[
\hat{S}^{(2)}(\cdot) = \nu^{(2)}\left( \left\{ x \in \mathbb{E}^{(2)} : x^{(2)} \geq 1, \frac{x^2}{x^1 + x^2} \in \cdot \right\} \right). \quad (4.7.2)
\]
Figure 4.2: Hill plot of $\alpha^{(2)}$ and estimated and actual transformed hidden angular densities

The density with respect to Lebesgue measure of $\tilde{S}^{(2)}(\cdot)$ is

$$f(s) = \begin{cases} \frac{1}{2}(1 - s)^{-2}, & \text{if } 0 \leq s < \frac{1}{2}, \\ \frac{1}{2} s^{-2}, & \text{if } \frac{1}{2} \leq s \leq 1, \\ 0 & \text{otherwise}. \end{cases} \quad (4.7.3)$$

We test accuracy of our estimates of $\alpha^{(2)}$ and $\tilde{S}^{(2)}(\cdot)$. The Hill plot for $\{m_i^{(2)}, i = 1, 2, \cdots, n\}$, the plot of the estimated transformed hidden angular densities for
k = 500, 1000, and the plot of the actual transformed hidden angular density (4.7.3) are shown in Figure 4.2. We also estimate probabilities of risk sets of the form \( P[X_1 > t_1, Y_1 > t_2] \) for large thresholds \( t_1 \) and \( t_2 \). Using a method similar to the one used to obtain (4.6.10), we estimate the probability \( P[X_1 > t_1, Y_1 > t_2] \) as

\[
P[X_1 > t_1, Y_1 > t_2] \approx \frac{k}{n} \int_{\mathbb{R}^2} \left( \left[ t_1 \left( \frac{Y}{X} \right)^{\tilde{\beta}_1} \right] \vee \left[ t_2 \left( \frac{Y}{X} \right)^{\tilde{\beta}_2} \right] \right)^{-\hat{\nu}^{(2)}} (\hat{S}(d\theta),
\]

where \( X_1 \geq X_2 \geq \cdots \geq X_n \) and \( Y_1 \geq Y_2 \geq \cdots \geq Y_n \) are the order statistics for \( \{X_i, i = 1, \cdots, n\} \) and \( \{Y_i, i = 1, \cdots, n\} \), and the remaining notation has the same meaning as in (4.6.10). Since we simulate the data we take \( \hat{\beta}_1 = 1 \) and \( \hat{\beta}_2 = 2 \) and concentrate on estimating \( \alpha^{(2)} \) using the Hill estimator and estimating \( S^{(2)}(\cdot) \) by the formula given in (4.3.14). We compute the estimates of \( P[X_1 > 100, Y_1 > \sqrt{10}] \) for different values of \( k \) using these estimators and plot the graph in Figure 4.3. The range of \( k \) is \( k = 500 \) to \( k = 5000 \).

A different estimate of \( P[X_1 > t_1, Y_1 > t_2] \) is obtained following Heffernan and Resnick [27]:

\[
P[X_1 > t_1, Y_1 > t_2] \approx \frac{k}{n} \hat{\nu}^{(2)} \left( \left[ \frac{t_1}{X} \left( \frac{Y}{X} \right)^{\tilde{\beta}_1} \right], \left[ \frac{t_2}{Y} \left( \frac{Y}{X} \right)^{\tilde{\beta}_2} \right], \infty \right),
\]

where \( \hat{\nu}^{(2)}(\cdot) \) is defined in (4.3.11). We again use \( \hat{\beta}_1 = 1 \), \( \hat{\beta}_2 = 2 \) and estimate \( \alpha^{(2)} \) using the Hill estimator. Then, using the above estimator, we compute the probability \( P[X_1 > 100, Y_1 > \sqrt{10}] \) for different values of \( k \) and plot it as a graph in Figure 4.3. The values of \( k \) are chosen between \( k = 500 \) and \( k = 5000 \).

Using the true distribution of \( (X_1, Y_1) \), we calculate \( P[X_1 > 100, Y_1 > \sqrt{10}] = 0.001 \). In Figure 4.3, we observe that the plot of the risk estimates obtained using the Heffernan-Resnick [27] estimator is more stable but our current estimator of \( P[X_1 > 100, Y_1 > \sqrt{10}] \) is more accurate for most \( k \) in the range \( k = 500 \) to \( k = 5000 \).
Figure 4.3: Plots of estimates of $P[X_1 > 100, Y_1 > \sqrt{10}]$ for different values of $k$ (sample size = 5000) using both our and H-R(Heffernan-Resnick [27]) estimator.

The Heffernan-Resnick [27] estimator of $P[X_1 > t_1, Y_1 > t_2]$ uses an empirical distribution function and thus is subject to the defect that a zero estimate is reported for the risk probability when $t_1$ and $t_2$ are high but actually $P[X_1 > t_1, Y_1 > t_2]$ is non-zero. Irrespective of how high the threshold is, our estimator does not estimate $P[X_1 > t_1, Y_1 > t_2]$ as zero, unless it is actually zero.

As an illustration, we reduced the sample size to $n = 500$ and applied the two estimators of the risk probability $P[X_1 > 100, Y_1 > \sqrt{10}]$. As suspected, the Heffernan-Resnick [27] estimator estimates the probability $P[X_1 > 100, Y_1 > \sqrt{10}]$ as zero, whereas our estimator is still reasonably accurate. This is shown in Figure 4.4 where $k$ ranges between $k = 50$ and $k = 500$. 
4.7.2 Internet traffic data

We analyze HTTP Internet response data consisting of sizes and durations of responses collected during a four hour period from 1–5 pm on April 26, 2001 by the University of North Carolina at Chapel Hill Department of Computer Science’s Distributed and Real-Time Systems Group under the direction of Don Smith and Kevin Jaffey. This dataset was also analyzed in [27]. We investigate joint behavior of two variables - size of response and rate (size of response/time duration of response) and estimate the probability that both the size and rate are big as a measure of burstiness.

We start by estimating marginal tail parameters. We use QQ plots [47, page 97] (not shown here) to choose the value $k = 5000$ for both the variables size and rate. Using this $k$, we get the estimates of tail indices $\hat{\beta}_1 = 1.15$ and $\hat{\beta}_2 = 1.51$ for size and rate using the QQ estimator.
Next, we investigate presence of asymptotic independence by plotting an estimated density of the transformed angular measure $\tilde{S}(\cdot)$, defined in Section 4.4.1. In agreement with Heffernan and Resnick [27], our estimated density plots for different values of $k$ for the transformed angular measure show two modes at the points 0 and 1 and take values close to zero in between, thus indicating asymptotic independence of size and rate (plots are not shown).

Is hidden regular variation present? The Hill plot in Figure 4.5 of $\{m_i^{(2)}, 1 \leq i \leq n\}$ suggests this is so and we proceed to estimate the density of the transformed hidden angular measure. Figure 4.5 gives plots of the estimated transformed hidden angular densities for $k = 500, 1000, 5000$.

Next, we estimate probabilities of risk sets of the form $[\text{Size} > x, \text{Rate} > y]$ which we consider as measures of burstiness. Examination of the (Size, Rate) data, indicates $x = 2 \times 10^7$ and $y = 10^5$ are reasonably high thresholds. We use both our estimator and the estimator given in [27] to compute $P[\text{Size} > x, \text{Rate} > y]$ for different values of $k$ from $k = 500$ to $k = n$ and plot them in Figure 4.6.

We also estimated $P[\text{Size} > x, \text{Rate} > y]$ for higher thresholds $x = 2 \times 10^8$ and $y = 10^7$, as a measure of extreme traffic burstiness. Again, we use both our estimator and the Heffernan-Resnick [27] estimator to estimate $P[\text{Size} > x, \text{Rate} > y]$ for different values of $k$ from $k = 500$ to $k = n$ and plot them in Figure 4.7. If hidden regular variation is present for the pair (Size, Rate), then the actual risk probability cannot be zero. The Heffernan-Resnick [27] estimator reports an estimate of zero but ours does not.
Figure 4.5: Hill plot of $a^{(2)}$ and estimated transformed hidden angular densities

4.8 Concluding remarks

Hidden regular variation provides a sub-family of the distributions having regular variation on $\mathbb{E}$ that is sometimes equipped to obtain more precise estimates of probabilities of certain risk sets, which are crudely estimated as zero by regular variation on $\mathbb{E}$; two examples are shown in Section 4.6.
Figure 4.6: Plots of estimates of $P[\text{Size} > 2 \times 10^7, \text{Rate} > 10^5]$ for different values of $k$ using our estimator and H-R(Heffernan-Resnick [27]) estimator

Figure 4.7: Plots of estimates of $P[\text{Size} > 2 \times 10^8, \text{Rate} > 10^6]$ for different values of $k$ using our estimator and H-R(Heffernan-Resnick [27]) estimator
The theory of HRV has deficiencies. Consider \( d = 3 \), and on the planes of \( \mathbb{E}^{(2)} \), suppose the random vector \( Z \) has regular variation with three different tail indices \( \alpha^{(2),1} < \alpha^{(2),2} < \alpha^{(2),3} \). As a convention, say \( Z \) has regular variation with tail index \( \alpha^{(2),i} \) on \( \{ x \in \mathbb{E}^{(2)} : x^i = 0 \} \), \( i = 1, 2, 3 \). If we follow our HRV model and method of estimation, we ignore the regular variation \( Z \) exhibits on \( \{ x \in \mathbb{E}^{(2)} : x^2 = 0 \} \) with tail index \( \alpha^{(2),2} \), which is actually more important than the regular variation on \( \mathbb{E}^{(3)} \). A way to repair this defect is the following alternative method: In \( d \) dimensions, first consider the big cone \( \mathbb{E} \), then consider all the \( \binom{d}{2} \) pairs of components of \( Z \) and their regular variation on \( (0, \infty)^2 \), then consider all the \( \binom{d}{3} \) triplets of components of \( Z \) and their regular variation on \( (0, \infty)^3 \), and so on. This alternative method requires considering regular variation on \( 2^d - 1 \) cones, whereas our HRV formulation requires considering regular variation on at most \( d \) cones. The alternative method is difficult to apply in high dimensions. Obviously, there is considerable flexibility in choosing a nested sequence of cones and informed choice by a practitioner will be governed by the application.

Another potential defect of our formulation of HRV is that it is designed to deal with the kind of degeneracy which arises when the limit measures are concentrated on the axes, planes etc. But the limit measures might exhibit different kind of degeneracies. For example, consider the degeneracy in the case of complete asymptotic dependence, where the limit measure is concentrated on the ray \( \{ x \in \mathbb{E} : x^1 = x^2 = \cdots = x^d \} \). One might think of removing the ray and considering hidden regular variation on the cone \( \mathbb{E} \setminus \{ x \in \mathbb{E} : x^1 = x^2 = \cdots = x^d \} \). Our HRV discussion does not address this issue.

Other variants of our formulation are possible. The hidden variation on a
sub-cone could be of extreme value type other than regular variation and even if we focus only on the hidden variation being regular variation, one could envisage different scaling functions for the hidden variation.

In estimating the limit measure $\nu^{(l)}(\cdot)$ of hidden regular variation on $\mathbb{E}^{(l)}$, $2 \leq l \leq d$, we have suggested a method that exploits the semi-parametric structure of $\nu^{(l)}(\cdot)$. Also, we have constructed a consistent estimator of $\nu^{(l)}(\cdot)$ which relies completely on non-parametric methods, as given in (4.3.10). Our numerical experiments in Section 4.7 clearly suggest the method exploiting the semi-parametric structure is superior, presumably because it uses more available information about the limit measure $\nu^{(l)}(\cdot)$. However, we have no precise, provable comparison.

An important statistical issue is we have only developed parameter estimators which are consistent. We have not yet developed theory which allows one to report on confidence intervals for parameter estimates or risk probability estimates.

For characterizations of hidden regular variation, it is important to identify when $\nu^{(l)}(\{x \in \mathbb{E}^{(l)} : \|x\| > 1\})$ is finite. We found a moment condition to check this, but it requires knowledge of the hidden angular measure $S^{(l)}(\cdot)$. A similar problem appeared in checking the finiteness of the right side of (4.6.12). It would be useful to have a statistical test for finiteness.
4.9 Proofs

Proof of Proposition 4.3.1. The idea of the proof is similar to Proposition 2 of [46]. Define \( E_{\infty} = E(l) \cup \bigcup_{1 \leq i_1 < i_2 < \cdots < i_d} \{ x^{i_1} = \infty, x^{i_2} = \infty, \ldots, x^{i_d} = \infty \} \) and \( E_2 = (0, \infty) \times \delta N^{(l)} \).

Define a continuous bijection \( Q^{(l)} : E_{\infty} \mapsto E_2 \) as in (4.3.4). We first show the equivalence of the vague convergence of measures restricted to \( E_{\infty} \) and \( E_2 \), and then extend the convergence to the corresponding whole spaces using the scaling property.

**Step 1:** First, we prove the direct part. So, we suppose that (4.2.10) holds with \( \nu^{(l)}(N^{(l)}) = 1 \). Hence, the convergence also holds with the measures being restricted to \( E_{\infty} \), that is

\[
\begin{aligned}
\mathbb{P} \left[ \frac{Z}{b^{(l)}(t)} \in \cdot \cap E_{\infty} \right] \xrightarrow{\nu} \nu^{(l)}(\cdot \cap E_{\infty}) \quad \text{in } M_+(E_{\infty}).
\end{aligned}
\]

Now, we proceed to show that for each compact set \( K_2 \) in \( E_2 \), \( (Q^{(l)})^{-1}(K_2) \) is a compact set of \( E_{\infty} \). Note that the compact sets in \( E_{\infty} \) are those closed sets \( K \) for which every \( x \in K \) satisfies the property that \( r \leq x^{(l)} \leq s \) for some \( 0 < r < s \) [47, page 170]. Take a compact set \( K_2 \) in \( E_2 \). We claim that \( K_2 \) must be contained in a set \( \tilde{K}_2 \) of the form \( \tilde{K}_2 = [r, s] \times \delta N^{(l)} \). Now, from the description of the compact sets of \( E_{\infty} \), \( (Q^{(l)})^{-1}(\tilde{K}_2) = \{ x \in E^{(l)} : r \leq x^{(l)} \leq s \} \) is compact in \( E_{\infty} \). Also, since \( Q^{(l)}(\cdot) \) is continuous, \( (Q^{(l)})^{-1}(K_2) \) is closed. Therefore, \( (Q^{(l)})^{-1}(K_2) \) is a closed subset of the compact set \( (Q^{(l)})^{-1}(\tilde{K}_2) \) and hence is compact in \( E_{\infty} \). So, using Proposition 5.5 (b) of [47] we get

\[
\begin{aligned}
\mathbb{P} \left[ \left( \frac{Z^{(l)}}{b^{(l)}(t)}, \frac{Z}{Z^{(l)}} \right) \in \cdot \cap E_2 \right] \xrightarrow{\nu} \nu^{(l)}(\cdot \cap E_2) \times S^{(l)}(\cdot \cap E_2) \quad \text{in } M_+(E_2).
\end{aligned}
\]

Now, we want to extend the convergence over the whole space \( (0, \infty) \times \delta N^{(l)} \). Choose any relatively compact subset \( \Lambda \) of \( \delta N^{(l)} \) such that \( S^{(l)}(\delta \Lambda) = 0 \) and choose
s > r > 0. Then,
\[
t_P \left[ \frac{Z(l)}{b(l)(t)} > r, \frac{Z}{Z(l)} \in \Lambda \right] \geq t_P \left[ \frac{Z(l)}{b(l)(t)} \in (r, s], \frac{Z}{Z(l)} \in \Lambda \right] \to \nu_{\alpha(l)}((r, s])S^{(l)}(\Lambda)
\]
as \(t \to \infty\), which implies that for \(s > r > 0\),
\[
\lim \inf_{t \to \infty} t_P \left[ \frac{Z(l)}{b(l)(t)} > r, \frac{Z}{Z(l)} \in \Lambda \right] \geq \nu_{\alpha(l)}((r, s])S^{(l)}(\Lambda).
\]
Hence, letting \(s \to \infty\), we get
\[
\lim \inf_{t \to \infty} t_P \left[ \frac{Z(l)}{b(l)(t)} > r, \frac{Z}{Z(l)} \in \Lambda \right] \geq r - \alpha(l)S^{(l)}(\Lambda).
\]
(4.9.1)

Now, we know that
\[
t_P \left[ \frac{Z(l)}{b(l)(t)} > r, \frac{Z}{Z(l)} \in \Lambda \right] = t_P \left[ \frac{Z(l)}{b(l)(t)} \in (r, s], \frac{Z}{Z(l)} \in \Lambda \right] + t_P \left[ \frac{Z(l)}{b(l)(t)} > s, \frac{Z}{Z(l)} \in \Lambda \right],
\]
and
\[
\lim \lim \sup_{s \to \infty} t_P \left[ \frac{Z(l)}{b(l)(t)} > s, \frac{Z}{Z(l)} \in \Lambda \right] \leq \lim \sup_{s \to \infty} t_P \left[ \frac{Z(l)}{b(l)(t)} > s \right]
\]
\[
= \lim s^{-\alpha(l)}\nu(l)(\{x \in \mathbb{R}^{(l)} : x(l) > s\}) = \lim s^{-\alpha(l)}\nu(l)(\{x \in \mathbb{R}^{(l)} : x(l) > 1\}) = 0.
\]
(4.9.3)
The first equality in the above set of relations follows from (4.2.10). Hence, from (4.9.2) and (4.9.3), we get
\[
\lim \sup_{t \to \infty} t_P \left[ \frac{Z(l)}{b(l)(t)} > r, \frac{Z}{Z(l)} \in \Lambda \right] \leq \lim \sup_{s \to \infty} t_P \left[ \frac{Z(l)}{b(l)(t)} \in (r, s], \frac{Z}{Z(l)} \in \Lambda \right] = r^{-\alpha(l)}S^{(l)}(\Lambda).
\]
(4.9.4)
Hence, from (4.9.1) and (4.9.4), we conclude the direct part of the proof:
\[
\lim t_P \left[ \frac{Z(l)}{b(l)(t)} > r, \frac{Z}{Z(l)} \in \Lambda \right] = r^{-\alpha(l)}S^{(l)}(\Lambda).
\]
Step 2: To see the converse, again we prove first the vague convergence of the restricted measures in \( M_+(\mathbb{E}_{\uparrow \omega}) \) and then extend it to convergence of measures in \( M_+(\mathbb{E}^l) \). We assume that (4.3.1) holds. Restriction on \( \mathbb{E}_2 \) gives

\[
\nu^l(b(t)) \rightarrow \nu_{\nu^l} \times S^l(\cdot \cap \mathbb{E}_2) \quad \text{in} \ M_+(\mathbb{E}_2).
\]

First we note that the compact sets of \( \mathbb{E}_2 \) are those closed sets \( K \) for which every \((w, v) \in K\) satisfies the property that \( r \leq w \leq s \) for some \( 0 < r < s \). Take a compact set \( K_1 \) of \( \mathbb{E}_{\uparrow \omega} \). Observe, from the description of the compact sets of \( \mathbb{E}_{\uparrow \omega} \) as given before, that \( K_1 \) must be contained in a set \( \bar{K}_1 \) of the form \( \bar{K}_1 = \{ x \in \mathbb{E}^l : x^l \in [r, s] \} \). From the description of compact sets of \( \mathbb{E}_2 \), \( Q^l(\bar{K}_1) = [r, s] \times \mathbb{N} \) is compact in \( \mathbb{E}_2 \). Since \((Q^l)^{-1}(-)\) is also continuous, the set \( Q^l(K_1) \) is closed, and hence, being a closed subset of a compact set \( Q^l(\bar{K}_1) \), is compact. Therefore, using the continuous map \((Q^l)^{-1} : \mathbb{E}_2 \rightarrow \mathbb{E}_{\uparrow \omega} \) and Proposition 5.5 (b) of [47], we get

\[
tP \left[ \frac{Z^l}{b^l(t)} \in \mathbb{E}_2 \right] \rightarrow \nu^l(\cdot \cap \mathbb{E}_{\uparrow \omega}) \quad \text{in} \ M_+(\mathbb{E}_{\uparrow \omega}).
\]

Now, we want to extend this convergence over the whole space \( \mathbb{E}^l \). Choose a relatively compact set \( A \) of \( \mathbb{E}^l \) such that \( \nu^l(\delta A) = 0 \). Note that, from the description of relatively compact sets in \( \mathbb{E}^l \) as given in Section 4.2.2, \( A \subseteq \{ x \in \mathbb{E}^l : x^l > r \} \) for some \( r > 0 \). Also, from the earlier description of compact sets of \( \mathbb{E}_{\uparrow \omega} \), it follows that for all \( s > r \), \( A \cap \{ x \in \mathbb{E}^l : x^l < s \} \) is a relatively compact set of \( \mathbb{E}_{\uparrow \omega} \), and

\[
\nu^l(\delta(A \cap \{ x \in \mathbb{E}^l : x^l < s \})) \leq \nu^l(\delta A) + \nu^l(\{ x \in \mathbb{E}^l : x^l = s \}) = 0.
\]

Hence, it follows that

\[
tP \left[ \frac{Z}{b^l(t)} \in A \right] \geq tP \left[ \frac{Z}{b^l(t)} \in A \cap \{ x \in \mathbb{E}^l : x^l < s \} \right] \rightarrow \nu^l(A \cap \{ x \in \mathbb{E}^l : x^l < s \}).
\]
which implies, by letting $s \to \infty$,

$$
\liminf_{t \to \infty} tP \left[ \frac{Z}{b^{(l)}(t)} \in A \right] \geq \nu^{(l)}(A),
$$

(4.9.5)
since by (4.2.11), $\nu^{(l)}(\{x \in \mathbb{B}^{(l)} : x^{(l)} = \infty\}) = 0$. Now, we know that

$$
tP \left[ \frac{Z}{b^{(l)}(t)} \in A \right] = tP \left[ \frac{Z}{b^{(l)}(t)} \in A \cap \{x \in \mathbb{B}^{(l)} : x^{(l)} < s\} \right] + tP \left[ \frac{Z}{b^{(l)}(t)} \in A \cap \{x \in \mathbb{B}^{(l)} : x^{(l)} \geq s\} \right],
$$

(4.9.6)
and

$$
\limsup_{s \to \infty} \limsup_{t \to \infty} tP \left[ \frac{Z}{b^{(l)}(t)} \in A \cap \{x \in \mathbb{B}^{(l)} : x^{(l)} \geq s\} \right] \\
\leq \limsup_{s \to \infty} \limsup_{t \to \infty} tP \left[ \frac{Z}{b^{(l)}(t)} \in \{x \in \mathbb{B}^{(l)} : x^{(l)} \geq s\} \right] \\
= \limsup_{s \to \infty} \limsup_{t \to \infty} \left[ \frac{Z^{(l)}(b)}{b^{(l)}(t)} \geq s \right] = \limsup_{s \to \infty} \nu_{a^{(l)}}(s, \infty) = \limsup_{s \to \infty} s^{-a^{(l)}} = 0.
$$

(4.9.7)
The second equality in the above set of relations follows from (4.3.1). Hence, from (4.9.6) and (4.9.7), we get

$$
\limsup_{t \to \infty} tP \left[ \frac{Z}{b^{(l)}(t)} \in A \cap \{x \in \mathbb{B}^{(l)} : x^{(l)} < s\} \right] = \nu^{(l)}(A).
$$

(4.9.8)
Therefore, from (4.9.5) and (4.9.8), we conclude

$$
\lim_{t \to \infty} tP \left[ \frac{Z}{b^{(l)}(t)} \in A \right] = \nu^{(l)}(A),
$$
which completes the converse part of the proof.

□

Proof of Proposition 4.3.7. The idea of this proof is similar to Theorem 6.1 of [47, page 173]. Define $\mathbb{B}_{l,\infty} = \mathbb{B}^{(l)} \setminus \cup_{1 \leq i_1 < \cdots < i_l \leq d} [x^{i_1} = \infty, x^{i_2} = \infty, \cdots, x^{i_l} = \infty]$ and $\mathbb{B}_2 = (0, \infty) \times \delta \mathbb{N}$. Now, define the continuous bijection $Q^{(l)} : \mathbb{B}_{l,\infty} \leftrightarrow \mathbb{B}_2$ as in
(4.3.4). As in Proposition 4.3.1, here also the idea of the proof is to first show the equivalence of the weak convergence of random measures restricted to \( M_+ (\mathbb{E}_{l,\infty}) \) and \( M_+ (\mathbb{E}_2) \), and then extend the convergence to the corresponding whole spaces using the scaling property.

Step 1: First, we prove that (4.3.11) implies (4.3.12). The convergence in (4.3.11) implies

\[ \hat{\nu}^{(l)} (\cdot \cap \mathbb{E}_{l,\infty}) := \frac{1}{k} \sum_{i=1}^{n} \epsilon (1/r_i^l / m_i^{(l)}), (1/\ell_i^l / m_i^{(l)}), 1 \leq j \leq d) 1_{[m_i^{(l)} / m_i^{(l)} < \infty]} \Rightarrow \nu^{(l)} (\cdot \cap \mathbb{E}_{l,\infty}) \]

on \( M_+ (\mathbb{E}_{l,\infty}) \). Also, as shown in the proof of Proposition 4.3.1, for any compact set \( K_2 \subset \mathbb{E}_2 \), \((Q^{(l)})^{-1}(K_2)\) is compact in \( \mathbb{E}_{l,\infty} \). Then, using Proposition 5.5(b) of [47] we get

\[ \hat{\nu}^{(l)} (\cdot \cap \mathbb{E}_{l,\infty}) \circ (Q^{(l)})^{-1} = \frac{1}{k} \sum_{i=1}^{n} \epsilon (m_i^{(l)} / m_i^{(l)}), (1/\ell_i^l / m_i^{(l)}), 1 \leq j \leq d) 1_{[m_i^{(l)} / m_i^{(l)} < \infty]} \Rightarrow \nu^{(l)} \times S^{(l)}(\cdot \cap \mathbb{E}_2) \]

on \( M_+ (\mathbb{E}_2) \). To extend the convergence to the space \((0, \infty) \times \delta \mathbb{N}^{(l)}\), we use the convergence of laplace functionals and use Theorem 5.2 of [47, page 137]. Take \( f(\cdot) \in C^+_K((0, \infty) \times \delta \mathbb{N}^{(l)}) \), where \( C^+_K(\mathbb{F}) \) is the set of all continuous functions with compact support from \( \mathbb{F} \) to \( \mathbb{R}^+ \). To relate this function to one defined in \( C^+_K((0, \infty) \times \delta \mathbb{N}^{(l)}) \), for all \( \delta, M > 0 \), we define a truncation function

\[ \phi_{\delta,M}(t) = \begin{cases} 1 & \text{if } 0 < t \leq M \\ 0 & \text{if } t > M + \delta, \\ \text{linear interpolation} & \text{if } M < t \leq M + \delta. \end{cases} \]

Note that \( f_{\delta,M}(r, \theta) := f(r, \theta) \phi_{\delta,M}(r) \in C^+_K((0, \infty) \times \delta \mathbb{N}^{(l)}) \) for all \( \delta, M > 0 \). Note that

\[
|E \left[ \exp \left[ -\frac{1}{k} \sum_{i=1}^{n} f \left( m_i^{(l)} / m_i^{(l)}(h) \right), (1/r_i^l / m_i^{(l)}), 1 \leq j \leq d \right) \right] - \exp \left[ -\nu^{(l)} \times S^{(l)}(f) \right] | \\
\leq |E \left[ \exp \left[ -\frac{1}{k} \sum_{i=1}^{n} f \left( m_i^{(l)} / m_i^{(l)}(h) \right), (1/r_i^l / m_i^{(l)}), 1 \leq j \leq d \right) \right] |
\]

164
\[-E \left[ \exp \left( -\frac{1}{k} \sum_{i=1}^{n} f_{\delta,M} \left( m_i^{(l)}/m_{i(k)}^{(l)} \right), \left( (1/r_i^l)/m_i^{(l)}, 1 \leq j \leq d \right) \right) \right] + \left| E \left[ \exp \left( -\frac{1}{k} \sum_{i=1}^{n} f_{\delta,M} \left( m_i^{(l)}/m_{i(k)}^{(l)} \right), \left( (1/r_i^l)/m_i^{(l)}, 1 \leq j \leq d \right) \right) \right] \right] - \exp \left( -\nu \delta \times S^{(l)}(f_{\delta,M}) \right) + \left| \exp \left( -\nu \delta \times S^{(l)}(f_{\delta,M}) \right) - \exp \left( -\nu \delta \times S^{(l)}(f) \right) \right] = A + B + C.\]

Since, $f_{\delta,M} \in C^+_K((0, \infty) \times \delta N^{(l)})$, by (4.9.9), we get $\lim_{n \to \infty} B = 0$. Now, we proceed to show that $\lim_{M \to \infty} \limsup_{n \to \infty} A = 0$. Notice that

\[
\limsup_{n \to \infty} \left| E \left[ \exp \left( -\frac{1}{k} \sum_{i=1}^{n} f \left( m_i^{(l)}/m_{i(k)}^{(l)} \right), \left( (1/r_i^l)/m_i^{(l)}, 1 \leq j \leq d \right) \right) \right] \right] - E \left[ \exp \left( -\frac{1}{k} \sum_{i=1}^{n} f \left( m_i^{(l)}/m_{i(k)}^{(l)} \right), \left( (1/r_i^l)/m_i^{(l)}, 1 \leq j \leq d \right) \right) \right] = \limsup_{n \to \infty} \left\{ \exp \left( -\frac{1}{k} \sum_{i=1}^{n} f \left( m_i^{(l)}/m_{i(k)}^{(l)} \right), \left( (1/r_i^l)/m_i^{(l)}, 1 \leq j \leq d \right) \right) \right\} \times \left( 1 - \exp \left( -\frac{1}{k} \sum_{i=1}^{n} \left( f_{\delta,M} - f \right) \left( m_i^{(l)}/m_{i(k)}^{(l)} \right), \left( (1/r_i^l)/m_i^{(l)}, 1 \leq j \leq d \right) \right) \right) \right| \leq \limsup_{n \to \infty} \left\{ \left( 1 - \exp \left( -\frac{1}{k} \sum_{i=1}^{n} \left( f_{\delta,M} - f \right) \left( m_i^{(l)}/m_{i(k)}^{(l)} \right), \left( (1/r_i^l)/m_i^{(l)}, 1 \leq j \leq d \right) \right) \right) \right\},
\]

which, using the facts that $\|f\| = \sup_{(r,\theta) \in (0,\infty) \times \delta N^{(l)}} f(r, \theta) < \infty$, $\|f_{\delta,M} - f\| \leq \|f\| \cdot \|\phi_{\delta,M} - 1\| \leq \|f\|$ and $(f_{\delta,M} - f)(x, \theta) = 0$ for $x < M$, is bounded by

\[
\limsup_{n \to \infty} E \left[ 1 - \exp \left( -\frac{1}{k} \|f\| \sum_{i=1}^{n} \epsilon \left( m_i^{(l)}/m_{i(k)}^{(l)}, \left( (1/r_i^l)/m_i^{(l)}, 1 \leq j \leq d \right) \right) \right) \right] = E \left[ 1 - \exp \left( -\frac{1}{k} \|f\| \sum_{i=1}^{n} \epsilon \left( (1/r_i^l)/m_i^{(l)}, 1 \leq j \leq d \right) \right) \right],
\]

which, by (4.3.11), converges as $n \to \infty$, to

\[
1 - \exp \left( -\|f\| \cdot (x \in \mathbb{E}^{(l)} : x^{(l)} \in [M, \infty]) \right) = 1 - \exp \left( -\|f\| \cdot M^{-\alpha} \nu^{(l)}(N^{(l)}) \right)
\]

165
= 1 − \exp\left[−\|f\| \cdot M^{−\alpha}\right] \to 0,

as \( M \to \infty \). The argument for \( \lim_{M \to \infty} C = 0 \) is similar and is omitted. Hence, by Theorem 5.2 of [47, page 137], we obtain (4.3.12).

**Step 2:** To see the other part, that is (4.3.12) implies (4.3.11), we use a similar method. The convergence in (4.3.12) implies

\[
\frac{1}{k} \sum_{i=1}^{n} \epsilon((m_i^{(0)}/m_i^{(0)}, 1 \leq j \leq d))((\cdot) \cap \mathbb{B}) \Rightarrow \nu^{(0)} \times S((\cdot) \cap \mathbb{B})
\]

on \( M_+\mathbb{B} \). It is easy to see that \((Q^{(0)})^{-1}(\cdot)\) is a continuous bijection. Also, as shown in the proof of Proposition 4.3.1, for any compact set \( K_1 \subset \mathbb{B}_{(\infty)} \), \( Q^{(0)}(K_1) \) is compact in \( \mathbb{B} \). Therefore, using Proposition 5.5(b) of [47] we get

\[\hat{\nu}^{(0)}((\cdot) \cap \mathbb{B}_{(\infty)}) := \frac{1}{k} \sum_{i=1}^{n} \epsilon((1/r_i^{(0)})/m_i^{(0)}, 1 \leq j \leq d)1_{[m_i^{(0)}/m_i^{(0)} < \infty]} \Rightarrow \nu^{(0)}((\cdot) \cap \mathbb{B}_{(\infty)}) \tag{4.9.10}\]

on \( M_+\mathbb{B}_{(\infty)} \). We use the same truncation function \( \phi_{\delta,M}(\cdot) \) to relate functions on \( C_{K}^{+}(\mathbb{B}^{(0)}) \) to ones in \( C_{K}^{+}(\mathbb{B}_{(\infty)}) \). Choose \( f(\cdot) \in C_{K}^{+}(\mathbb{B}^{(0)}) \). Note that the function \( f_{\delta,M}(x) := f(x)\phi_{\delta,M}(x^{(0)}) \in C_{K}^{+}(\mathbb{B}_{(\infty)}) \) for all \( \delta, M > 0 \).

\[
|E\left[\exp\left[−\frac{1}{k} \sum_{i=1}^{n} f((1/r_i^{(0)})/m_i^{(0)}, 1 \leq j \leq d)\right]\right] – \exp[−\nu^{(0)}(f)]|
\leq |E\left[\exp\left[−\frac{1}{k} \sum_{i=1}^{n} f((1/r_i^{(0)})/m_i^{(0)}, 1 \leq j \leq d)\right]
\right.
\left.− \exp\left[−\frac{1}{k} \sum_{i=1}^{n} f_{\delta,M}\left((1/r_i^{(0)})/m_i^{(0)}, 1 \leq j \leq d\right)\right]\right|
\plus |E\left[\exp\left[−\frac{1}{k} \sum_{i=1}^{n} f_{\delta,M}\left((1/r_i^{(0)})/m_i^{(0)}, 1 \leq j \leq d\right)\right]− \exp[−\nu^{(0)}(f_{\delta,M})]\right|
\plus |\exp[−\nu^{(0)}(f_{\delta,M})]− \exp[−\nu^{(0)}(f)]|
= A + B + C.
\]
Since, $f_{\delta,M} \in C^+_K(\mathbb{E}_{l,\infty})$, by (4.9.10), we get $\lim_{n \to \infty} B = 0$. Now, we will show that

\[
\lim_{M \to \infty} \limsup_{n \to \infty} A = 0.
\]

\[
\limsup_{n \to \infty} \left| E \left[ \exp \left( -\frac{1}{k} \sum_{i=1}^{n} f \left( \frac{1}{r_j^i}/m_i^{(l)} \right), \ 1 \leq j \leq d \right) \right] - \exp \left( -\frac{1}{k} \sum_{i=1}^{n} f_{\delta,M} \left( \frac{1}{r_j^i}/m_i^{(l)} \right), \ 1 \leq j \leq d \right) \right| \right.
\]

\[
= \limsup_{n \to \infty} E \left[ \exp \left( -\frac{1}{k} \sum_{i=1}^{n} f \left( \frac{1}{r_j^i}/m_i^{(l)} \right), \ 1 \leq j \leq d \right) \right.
\]

\[
\times \left( 1 - \exp \left( -\frac{1}{k} \sum_{i=1}^{n} (f_{\delta,M} - f) \left( \frac{1}{r_j^i}/m_i^{(l)} \right), \ 1 \leq j \leq d \right) \right) \left. \right]
\]

\[
\leq \limsup_{n \to \infty} E \left[ \left( 1 - \exp \left( -\frac{1}{k} \sum_{i=1}^{n} (f_{\delta,M} - f) \left( \frac{1}{r_j^i}/m_i^{(l)} \right), \ 1 \leq j \leq d \right) \right) \right],
\]

which, using the facts $||f|| = \sup_{x \in \mathbb{E}^{(l)}} f(x) < \infty$, $||f_{\delta,M} - f|| \leq ||f|| \cdot ||\phi_{\delta,M} - 1|| \leq ||f||$ and $(f_{\delta,M} - f)(x) = 0$ for $\{x \in \mathbb{E}^{(l)} : x^{(l)} < M\}$, is bounded by

\[
E \left[ 1 - \exp \left( -\frac{1}{k} ||f|| \sum_{i=1}^{n} \epsilon_{(1/r_j^i)/m_i^{(l)}, 1 \leq j \leq d}(\{x \in \mathbb{E}^{(l)} : x^{(l)} \in [M, \infty]\}) \right) \right]
\]

\[
= \limsup_{n \to \infty} E \left[ 1 - \exp \left( -||f|| \frac{1}{k} \sum_{i=1}^{n} \epsilon_{m_i^{(l)}/m_i^{(l)}, [M, \infty]} \right) \right],
\]

which, by (4.3.12), converges as $n \to \infty$, to

\[
1 - \exp \left[ -||f|| \cdot \nu_{\alpha^{(l)}}([M, \infty]) \right] = 1 - \exp \left[ -||f|| \cdot M^{-\alpha^{(l)}} \right] \to 0,
\]

as $M \to \infty$. The argument for $\lim_{M \to \infty} C = 0$ is similar and is omitted. Hence, we obtain (4.3.11) and this completes the proof. \(\Box\)
APPENDIX A
SOME CONVERGENCE CONCEPTS AND DISTRIBUTIONAL RELATIONS

For easy reference, here we review the concepts of tail equivalence, subexponential distributions and vague convergence.

A.1 Tail equivalence

Two distributions $F(\cdot)$ and $G(\cdot)$ are tail equivalent if they have the same right endpoint meaning $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\} = \sup\{x \in \mathbb{R} : G(x) < 1\} = x_G$ and

$$\lim_{x \to x_F} \frac{G(x)}{F(x)} = \lim_{x \to x_F} \frac{(1 - G(x))}{(1 - F(x))} \in (0, \infty).$$

A standard reference is [45].

A.2 Subexponential distributions

A non-negative random variable $X$ is subexponential if for $X, Y \overset{i.i.d.}{\sim} F(\cdot)$,

$$\lim_{x \to \infty} \frac{P(X + Y > x)}{P(X > x)} = 2.$$

The class of subexponential distributions is denoted $S$. A standard reference is [23].
A.3 Vague convergence

Consider a locally compact topological space $\mathbb{E}$ with countable base; for convenience, think of $\mathbb{E}$ as a finite dimensional Euclidean space or $\mathbb{R}^d$. We assume that the space $\mathbb{E}$ comes with a $\sigma$-field generated by the open sets or equivalently, the rectangles of $\mathbb{E}$. We denote the non-negative Radon measures on Borel subsets of $\mathbb{E}$ as $M_+(\mathbb{E})$. A sequence of measures $\mu_n \in M_+(\mathbb{E})$ converges vaguely to $\mu \in M_+(\mathbb{E})$ (written $\mu_n \xrightarrow{v} \mu$) if for all continuous functions $f(\cdot)$ with compact support,

$$\mu_n(f) = \int_{\mathbb{E}} f(x) \mu_n(dx) \xrightarrow{n \to \infty} \int_{\mathbb{E}} f(x) \mu(dx) = \mu(f).$$

Standard references include [32, 45, 47].
APPENDIX B
MAXIMAL DOMAIN OF ATTRACTION AND REGULAR VARIATION

We now review the concept of maximal domain of attraction and regular variation. In accordance with our need, our definition of regular variation here is restricted to distributions of random vectors. A more general treatment of regularly varying functions can be found in [8, 47].

B.1 Maximal domain of attraction

A \( d \)-dimensional distribution function \( H(\cdot) \) of \( X = (X_1, X_2, \cdots, X^d) \) belongs to the maximal domain of attraction of a \( d \)-dimensional distribution function \( G(\cdot) \) if for all continuity points \( x = (x^1, x^2, \cdots, x^d) \) of \( G(\cdot) \),

\[
\lim_{n \to \infty} H^n(a^1_n x^1 + b^1_n, a^2_n x^2 + b^2_n, \cdots, a^d_n x^d + b^d_n) = G(x)
\]

and all the marginal distributions of \( G(\cdot) \) are non-degenerate distributions. In other words, the component-wise maxima of i.i.d. random vectors \( \{X_i = (X^1_i, X^2_i, \cdots, X^d_i) : i = 1, 2, \cdots, n \} \) with common distribution \( H(\cdot) \), after suitable centering and scaling, converges weakly to \( G(\cdot) \), that is

\[
\left( \frac{\vee_{i=1}^n X^1_i - b^1_n}{a^1_n}, \frac{\vee_{i=1}^n X^2_i - b^2_n}{a^2_n}, \cdots, \frac{\vee_{i=1}^n X^d_i - b^d_n}{a^d_n} \right) \Rightarrow G(\cdot).
\]

We write \( H(\cdot) \in MDA(G) \) or \( X \in MDA(G) \).

Standard references include [18, 45].
B.2 Regular variation

Let $M_+(\mathbb{E})$ denote the set of all non-negative Radon measures on Borel subsets of $\mathbb{E} := [0, \infty]^d \setminus \{0\}$ and $\overset{\rightharpoonup}{\longrightarrow}$ denote vague convergence. The distribution of $Z = (Z^1, Z^2, \cdots, Z^d)$ is standard regularly varying on $\mathbb{E}$ with limit measure $\nu(\cdot)$ if there exist a function $b(t) \uparrow \infty$ as $t \to \infty$ and a non-negative non-degenerate Radon measure $\nu(\cdot) \neq 0$ such that

$$tP \left[ \frac{Z}{b(t)} \in \cdot \right] \overset{\rightharpoonup}{\longrightarrow} \nu(\cdot) \text{ in } M_+(\mathbb{E}),$$

and all the marginal measures of $\nu(\cdot)$ are non-zero.

Non-standard regular variation may hold when (B.2.1) fails, but

$$tP \left[ \left( \frac{Z^j}{a^j(t)}, j = 1, 2, \cdots, d \right) \in \cdot \right] \overset{\rightharpoonup}{\longrightarrow} \mu(\cdot) \text{ in } M_+(\mathbb{E})$$

for some scaling functions $a^1(\cdot), a^2(\cdot), \cdots, a^d(\cdot)$ satisfying $a^j(t) \uparrow \infty$, where $\mu(\cdot)$ is a non-negative non-zero Radon measure on $\mathbb{E}$, such that all the marginal measures of $\mu(\cdot)$ are non-zero and non-degenerate.

Distributions possessing regular variation (standard or non-standard) are always in some maximal domain of attraction. Standard references include [8, 18, 45, 47].
BIBLIOGRAPHY


