Multivariate Joint Tail Modelling and Score Tests of Independence

Alexandra Ramos

Submitted for the Degree of
Doctor of Philosophy
from the
University of Surrey

Department of Mathematics and Statistics
School of Electronics and Physical Sciences
University of Surrey
Guildford, Surrey GU2 7XH, U.K.

December 2002

© Alexandra Ramos 2002
Summary

Probabilistic and statistical aspects of extremes of univariate processes have been extensively studied, and recent developments in extremes have focused on multivariate theory and its application. Multivariate extreme value theory encompasses two separate aspects: marginal features, which may be handled by standard univariate methods, and dependence features. Both will be examined in this study.

First we focus on testing independence in multivariate extremes. All existing score tests of independence in multivariate extremes have non-regular properties that arise due to violations of the usual regularity conditions of maximum likelihood. Some of these violations may be dealt with using standard techniques, for example when independence corresponds to a boundary point of the parameter space of the underlying model. However, another type of regularity violation, the infinite second moment of the score function, is more difficult to deal with and has important consequences for applications, resulting in score statistics with non-standard normalisation and poor rates of convergence. We propose a likelihood based approach that provides asymptotically normal score tests of independence with regular normalisation and rapid convergence. The resulting tests are straightforward to implement and are beneficial in practical situations with realistic amounts of data.

A fundamental issue in applied multivariate extreme value (MEV) analysis is modelling dependence within joint tail regions. The primary aim of the remainder of this thesis is to develop a pseudo-polar framework for modelling extremal dependence that extends the existing classical results for multivariate extremes to encompass asymptotically independent tails. Accordingly, a constructional procedure for obtaining parametric asymptotically independent joint tail models is developed. The practical application of this framework is analysed through applications to bivariate simulated and environmental data, and joint estimation of dependence and marginal parameters via likelihood methodology is detailed. Inference under our models is examined and tests of extremal asymptotic independence and asymmetry are derived which are useful for model selection. In contrast to the classical MEV approach, which concentrates on the distribution of the normalised componentwise maxima, our framework is based on modelling joint tails and focuses directly on the tail structure of the joint survivor function. Consequently, this framework provides significant extensions of both the theoretical and applicable tools of joint tail modelling. Analogous point process theory is developed and the classical componentwise maxima result for multivariate extremes is extended to the asymptotically independent case. Finally, methods for simulating from two of our bivariate parametric models are provided.

Key words: Multivariate extreme values, Score tests of independence, Likelihood ratio tests, Asymptotic independence, Joint tail dependence models, Point process, Simulation.
Acknowledgements

I would like to thank everyone who has helped me throughout these three years.

First and foremost, Dr. Anthony Ledford, who was my supervisor during the first two years of my PhD and extended his commitment to the third year, for his excellent assistance, encouragements and guidance. His enthusiasm and dedication have always motivated me and working with him has been a great pleasure. I am deeply grateful for his relentless support.

My thanks also go to my PhD’s third year supervisor, Prof. Trevor Sweeting, for interesting conversations and suggestions, in particular for all his help during the final preparations of this dissertation.

My husband Miguel has given me an extraordinary support throughout these three years. His help was beyond the simple understanding and patience as he often spared the time he needed for his own work to help me, either listening to me or even using his mathematical knowledge to give me suggestions for my work. His love has been a constant presence and it is the source of my strength.

Last but not least, I would like to thank my family and friends for their permanent support and love.

Financial support has been provided by Universidade do Porto, Fundação para a Ciência e Tecnologia of the Portuguese government and FSE. I would like to thank all these institutions.
## Contents

Summary ....................................................................................................................... ii  
Acknowledgements ...................................................................................................... iii  
Contents .......................................................................................................................... iv  
List of Figures ............................................................................................................. ix  
List of Tables ................................................................................................................ x  

1 Background  
1.1 Introduction .......................................................................................................... 1  
1.2 Structure of the thesis .......................................................................................... 2  
1.3 Univariate extreme value theory ................................................................... 3  
   1.3.1 Probability framework ........................................................................... 3  
   1.3.2 The classical limit laws ........................................................................ 4  
   1.3.3 The generalised extreme value distribution .................................... 5  
   1.3.4 Point process characterisation .............................................................. 5  
      1.3.4.1 Point processes ....................................................................... 6  
      1.3.4.2 A point process result ............................................................ 7  
   1.3.5 Threshold methods: the generalised Pareto distribution .............. 8  
   1.3.6 The r-largest order statistics method ................................................. 8  
   1.3.7 Stationary sequences .............................................................................. 9  
      1.3.7.1 Point process results for stationary sequences .................. 12  
      1.3.7.2 Non-stationary sequences .................................................... 13  
1.4 Multivariate extremes ....................................................................................... 13  
   1.4.1 Notation and limiting distribution .................................................... 13  
   1.4.2 A point process result for multivariate extremes ......................... 14  
   1.4.3 Multivariate extreme value distribution ....................................... 15  
   1.4.4 Examples ............................................................................................. 17  
1.5 Statistical modelling .......................................................................................... 18
1.5.1 Fitting the univariate models .............................................................. 19
1.5.2 Univariate statistical methods .............................................................. 19
  1.5.2.1 The peaks over threshold (POT) method ........................................ 20
  1.5.2.2 Point process approach .................................................................. 21
  1.5.2.3 Distribution-free methods ............................................................... 21
1.5.3 Multivariate statistical methods ........................................................... 22
  1.5.3.1 Multivariate point process approach .............................................. 23
  1.5.3.2 Multivariate threshold models ....................................................... 23
  1.5.3.3 Fitting the multivariate models ....................................................... 24
1.6 Dependence measures for extreme values ....................................................... 26
  1.6.1 The measures $\chi$ and $\overline{\chi}$ ...................................................... 27
  1.6.2 The coefficient of tail dependence $\eta$ and $\mathcal{L}(r)$ .......................... 28
  1.6.3 Relationship between $\eta$ and $\mathcal{L}$ and the dependence measures $\chi$
      and $\overline{\chi}$ ....................................................................................... 29
1.7 Examples of application areas ................................................................. 29

2 Assessment and comparison of the various estimation procedures 31
  2.1 Introduction ............................................................................................ 31
  2.2 Analysis of simulated data .................................................................... 33
    2.2.1 Simulated data .................................................................................. 33
    2.2.2 Results and analysis ........................................................................ 34
  2.3 Summary and conclusions .................................................................... 45

3 Score tests of independence in multivariate extreme values 46
  3.1 Introduction ............................................................................................ 47
  3.2 Score tests of independence ................................................................... 47
    3.2.1 Regularised score tests of independence ........................................ 50
    3.2.2 Evaluating the variances $\sigma_1^2$ and $\sigma_2^2$ .............................. 54
  3.3 Comparison between regular and non-regular score tests ...................... 55
  3.4 Further asymptotic properties of likelihood inference .......................... 63
  3.5 Alternative dependence models ......................................................... 64
  3.6 Conclusions ........................................................................................... 66
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>A pseudo-polar representation of asymptotic independence</td>
<td>68</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>68</td>
</tr>
<tr>
<td>4.2</td>
<td>Modelling dependence within bivariate joint tails</td>
<td>70</td>
</tr>
<tr>
<td>4.3</td>
<td>Examples of parametric models</td>
<td>73</td>
</tr>
<tr>
<td>4.4</td>
<td>Marginal properties of ( F_{ST} )</td>
<td>83</td>
</tr>
<tr>
<td>4.5</td>
<td>Standard bivariate extreme value (BEV) case</td>
<td>84</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Relationship between the measure densities ( h ) and ( h^* )</td>
<td>85</td>
</tr>
<tr>
<td>4.6</td>
<td>Particular case: ( \eta = 1 )</td>
<td>86</td>
</tr>
<tr>
<td>4.7</td>
<td>Conclusions</td>
<td>88</td>
</tr>
<tr>
<td>5</td>
<td>Asymptotically independent joint tail modelling in practice</td>
<td>90</td>
</tr>
<tr>
<td>5.1</td>
<td>The modelling framework</td>
<td>90</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Known unit Fréchet margins</td>
<td>90</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Unknown margins</td>
<td>92</td>
</tr>
<tr>
<td>5.2</td>
<td>Likelihood</td>
<td>93</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Known unit Fréchet margins</td>
<td>93</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Unknown margins</td>
<td>94</td>
</tr>
<tr>
<td>5.3</td>
<td>Inference and diagnostics</td>
<td>96</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Test for asymptotic independence</td>
<td>96</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Test for ray independence</td>
<td>96</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Test for asymmetry</td>
<td>97</td>
</tr>
<tr>
<td>5.3.4</td>
<td>Diagnostics</td>
<td>98</td>
</tr>
<tr>
<td>5.4</td>
<td>Application to simulated data</td>
<td>99</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Inference and results</td>
<td>101</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Extrapolation</td>
<td>106</td>
</tr>
<tr>
<td>5.5</td>
<td>Application to real data</td>
<td>109</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Inference and results</td>
<td>110</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Extrapolation</td>
<td>115</td>
</tr>
<tr>
<td>5.5.3</td>
<td>Comments</td>
<td>117</td>
</tr>
<tr>
<td>5.6</td>
<td>Conclusion</td>
<td>117</td>
</tr>
</tbody>
</table>
# Contents

## 6 Point process results for asymptotic independence

6.1 Introduction ................................................................. 119

6.2 Distribution of the componentwise maxima \((M_{X,n}, M_{Y,n})\) out of those points which are simultaneously large ........................................ 119

6.2.1 Obtaining \(h_\eta\) from \(V_\eta\) .............................................. 122

6.3 Distribution of the componentwise maxima \((M_{S,n}, M_{T,n})\) ......................................................... 123

6.4 Examples of \(V_\eta\) ............................................................ 124

6.5 Marginal properties .......................................................... 125

6.6 Comments .......................................................................... 126

## 7 Simulation methods

7.1 Simulation from the distribution \(F_{ST}\) ........................................ 128

7.1.1 Using the modified logistic model ....................................... 128

7.1.2 Using the modified asymmetric logistic model ..................... 131

7.2 Simulation from the limiting processes .................................... 133

7.2.1 Simulation from the bivariate logistic distribution function \(G_\eta\) .................................................. 133

7.2.2 Simulation from the bivariate asymmetric logistic distribution function \(G_\eta\) ............................................... 134

7.3 Comments .......................................................................... 136

## 8 Extensions and further work

8.1 Modelling dependence with partially observed data .................. 138

8.1.1 Example ............................................................................ 139

8.2 Modelling dependence within multivariate joint tails ................ 141

8.2.1 Example ............................................................................ 143

8.3 Further work .......................................................................... 144

## Bibliography

148
List of Figures

2 Assessment and comparison of the various estimation procedures 31

2.1 Diagnostic plots of the POT model fit for the Fréchet case 38
2.2 Diagnostic plots of the POT model fit for the exponential case 39
2.3 Diagnostic plots of the POT model fit for the normal case 40
2.4 Scale and shape parameters estimates against threshold in POT model analysis of unit Fréchet data 42
2.5 Scale and shape parameters estimates against threshold in POT model analysis of exponential data 43
2.6 Scale and shape parameters estimates against threshold in POT model analysis of normal data 44

3 Score tests of independence in multivariate extreme values 46

3.1 Histograms and normal QQ-plots of the normalised scores under independence 53
3.2 Plots of the observed score contributions from region \( R_{11} \) 54
3.3 Standard deviations of the modified scores with the corresponding approximations superimposed 56
3.4 Power functions of normalised score statistics based on the 95% asymptotic critical value 59
3.5 The loss of power that results from censoring region \( R_{11} \) 60
3.6 Power functions of normalised score statistics based on the 95% empirical critical values 61
3.7 Plot showing the actual and nominal sizes for normalised score statistics 62
3.8 Histograms and QQ-plots of simulated values of the likelihood ratio statistics 65
List of Figures

4 A pseudo-polar representation of asymptotic independence 68

4.1 Plots of the ray dependence function for the modified logistic model 75
4.2 Plots of the profile of the survivor function $\tilde{F}_{ST}$ for the modified logistic model 76
4.3 Plots of the ray dependence function for the modified asymmetric logistic model 80

5 Asymptotically independent joint tail modelling in practice 89

5.1 Plot of the four sets of simulated data with the selected thresholds included 99
5.2 Estimates of the dependence parameters by fitting the modified logistic model, together with diagnostic estimates of $\eta$, for a range of structure variable threshold probabilities for the simulated data 102
5.3 Plots of the fitted function $g^*(w)$ for the modified logistic model together with the diagnostic estimate of $g^*$ for the simulated data 104
5.4 Joint density estimates obtained using the new logistic model for the simulated data 106
5.5 Extrapolated joint tail contour curves of survivor function for the true model and for the fitted new logistic tail model for the simulated data 107
5.6 Scatter plots of the environmental data with the selected thresholds included 108
5.7 Estimates of the dependence parameters by fitting the modified asymmetric logistic model for a range of structure variable threshold probabilities for each of the environmental data sets 112
5.8 Diagnostic estimates of $g^*(w)$ with approximate pointwise 95% confidence intervals and fitted ray dependence functions estimates obtained from joint analyses using the full logistic and asymmetric logistic models for the environmental data sets 114
5.9 Joint density estimates obtained using the full asymmetric logistic model for the environmental data 115
5.10 Extrapolated joint tail contour curves of the full asymmetric logistic survivor model for the environmental data 115

7 Simulation methods 127

7.1 Simulated points from the bivariate logistic $F_{ST}$ 129
7.2 Simulated points from the bivariate asymmetric logistic $F_{ST}$ 131
7.3 Simulated points from the bivariate logistic distribution function $G_{\eta}$ 134
7.4 Simulated points from the bivariate asymmetric logistic distribution function $G_{\eta}$ 136
## List of Tables

### 2 Assessment and comparison of the various estimation procedures

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Limiting values of the POT shape and scale parameters for the Fréchet, exponential and normal distributions</td>
<td>34</td>
</tr>
<tr>
<td>2.2</td>
<td>Shape and scale parameter estimates with associated standard errors using unit Fréchet data</td>
<td>35</td>
</tr>
<tr>
<td>2.3</td>
<td>Shape and scale parameter estimates with associated standard errors using exponential data</td>
<td>36</td>
</tr>
<tr>
<td>2.4</td>
<td>Shape and scale parameter estimates with associated standard errors using normal data</td>
<td>36</td>
</tr>
</tbody>
</table>

### 3 Score tests of independence in multivariate extreme values

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Simulated and asymptotic critical values of the normalised score statistics</td>
<td>58</td>
</tr>
</tbody>
</table>

### 5 Asymptotically independent joint tail modelling in practice

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Dependence parameter estimates obtained from the diagnostic and model fits for the simulated data</td>
<td>101</td>
</tr>
<tr>
<td>5.2</td>
<td>Dependence parameter estimates obtained by fitting the modified logistic and asymmetric logistic models to the unit Fréchet transformed data and to the original environmental data</td>
<td>110</td>
</tr>
<tr>
<td>5.3</td>
<td>Marginal parameter estimates of the GPD shape and scale parameters for each margin of the environmental data sets obtained through separate and joint analyses</td>
<td>111</td>
</tr>
</tbody>
</table>

### 8 Extensions and further work

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1</td>
<td>Dependence parameter estimates obtained by fitting the logistic univariate model and the logistic joint tail model using the chosen threshold ( u ) for the simulated data</td>
<td>139</td>
</tr>
</tbody>
</table>
Chapter 1

Background

1.1 Introduction

Extreme value theory concerns the statistical study of the extremal properties of random processes. This branch of statistics has primarily found applications in modelling environmental extremal phenomena, which may lead to substantial damage to property and impact on people's lives. Consequently, developments in extreme value theory have been motivated through applications, such as hydrology, offshore engineering or strength of materials. A typical example of its application is in the design of coastal defences, where the appropriate height of a sea dyke is decided by applying extreme value methods so that the dyke affords a certain degree of safety, usually specified by legislation or determined by cost. In general, extreme value theory methods provide a means of quantifying the frequency and intensity of the extremes of a random process.

The most common problems treated by extreme value methods involve modelling the tail of an unknown distribution function from a set of observed data with the purpose of quantifying the frequency and severity of events more extreme than any that have been observed previously. Since extreme data are, by definition, scarce there are often practical difficulties to be overcome due to sparse information. Extrapolation of the model beyond the range of observed data is required, a feature that distinguishes extreme value statistics from the majority of other areas of statistical modelling. Clearly, such applications require procedures which are scientifically and statistically rational for estimating the extremal behaviour of random variables or processes.

For univariate processes, both the probabilistic and statistical aspects of extremes have
been extensively studied, and recent developments in extremes have focused on multivariate theory and its applications. The multivariate extreme value theory results have two separate aspects: marginal features, which may be handled by standard univariate methods, and dependence features. We will examine both within this study.

1.2 Structure of the thesis

We now describe how the thesis chapters are organized.

The remainder of this chapter reviews the main results from the extreme value theory literature. First, results from univariate and multivariate extreme value theory are stated, and an outline of their statistical implementation is given. Several of the commonly used dependence measures for extreme values are then examined, and we briefly report some of the application areas of these results.

Chapter 2 presents illustrative applications showing how univariate extreme value methods may be implemented using applied statistical techniques such as maximum likelihood estimation, moment based estimation and Markov chain Monte Carlo. The performances of these estimation methods are assessed and compared using simulated data.

In Chapter 3 we propose score based tests with regular properties for testing independence of bivariate extreme values. These tests overcome many of the difficulties that arise due to the non-regularities encountered when using existing score tests of independence. Likelihood ratio tests and their properties are also studied.

A pseudo-polar representation of asymptotic independence in terms of a non-negative angular measure that is essentially arbitrary apart from having to satisfy a simple normalisation condition is given in Chapter 4. Asymptotically independent joint tail parametric models are then obtained from a constructional procedure provided by that representation. Practical applications of these models are examined in Chapter 5. There, modifications of the standard logistic and asymmetric logistic models are investigated using simulated and environmental data. Inference techniques are also studied.

In Chapter 6 point process results for asymptotic independence are studied and the classical componentwise maxima result for bivariate extremes is extended to the asymptotically independent case. Chapter 7 provides methods for simulating from the modified
symmetric and asymmetric bivariate logistic models. First, simulation from the joint tail models is considered and then simulation from the limit distribution of componentwise maxima of points that are simultaneously large is presented. Finally, in Chapter 8 further properties and advantages of our joint tail models are discussed and some future areas of research are suggested.

1.3 Univariate extreme value theory

In this section, we give an overview of some of the main results of univariate extreme value theory. These are stated without proof. For a detailed treatment of the results and their proofs see for example Leadbetter et al. (1983), Embrechts et al. (1997) or Coles (2000).

1.3.1 Probability framework

Let \( X_1, \ldots, X_n \) be a sequence of independent and identically distributed (iid) random variables with common distribution function \( F \). One simple way of characterising the behaviour of extremes is by considering the behaviour of the maximum \( M_n = \max(X_1, \ldots, X_n) \). Then, the distribution function of \( M_n \) is given by

\[
\Pr(M_n \leq x) = \Pr(X_1 \leq x, \ldots, X_n \leq x) = F^n(x)
\]

since the \( X \)'s are independent. For a fixed \( x \), this becomes degenerate (i.e. tends to 0 or 1) as \( n \to \infty \), since, with probability 1, \( M_n \) converges to the upper endpoint \( w(F) \) of the distribution function \( F \), defined by \( w(F) = \sup \{ x : F(x) < 1 \} \). Thus, as for central limit theory, some normalisation of \( M_n \) is required in order to obtain a non-degenerate limiting distribution. The usual form of normalisation that is chosen is a linear one\(^1\) and the aim is to determine the non-degenerate distributions \( G \) satisfying

\[
\Pr \left\{ \left( \frac{M_n - b_n}{a_n} \right) \leq x \right\} = F^n \left( a_n x + b_n \right) \converges \text{ at } G(x),
\]

(1.1)

for sequences \( a_n > 0 \) and \( b_n \in \mathbb{R} \), where \( \converges \) represents convergence at continuity points of the limiting function. Before stating the possible limiting distributions \( G \), it is convenient to define an equivalence class of distributions.

\(^1\)Non-linear normalisation is also possible but this is outside the standard theory.
Definition 1.1 (Type) Two distribution functions $G_1$ and $G_2$ are of the same type if $G_2(x) = G_1(ax + b)$ for some constants $a > 0$ and $b \in \mathbb{R}$.

1.3.2 The classical limit laws

The classical result for the iid case of univariate extreme value theory is the extremal types theorem of Fisher and Tippett (1928).

Theorem 1.2 (Extremal types theorem) If there exist sequences of constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that result (1.1) is satisfied for some non-degenerate distribution function $G$, then $G$ is of the same type as one of the following distributions:

- **Type 1:** $G(x) = \exp(-e^{-x})$, $-\infty < x < \infty$;
- **Type 2:**
  
  $G(x) = \begin{cases} 
  0 & x \leq 0, \\
  \exp(-x^{-\alpha}) & \text{for some } \alpha > 0, x > 0; \\
  \exp\{-(-x)^{\alpha}\} & \text{for some } \alpha > 0, x \leq 0, \\
  1 & x > 0.
  \end{cases}$

Conversely, each of these distribution functions $G$ may appear as a limit in (1.1), and, in fact, does so when $G$ itself is the distribution function of each $X_i$.

It is clear that if $F$ is of the same type as one of this three limiting distributions then the limiting distribution $G$ in (1.1) exists and is of the same type as $F$.

The three classes of distribution in Theorem 1.2 are referred to as the extreme value distributions, with types 1, 2 and 3 known as the Gumbel, Fréchet and Weibull types, respectively. Note that this theorem does not guarantee the existence of $G$ for an arbitrary $F$.

Another way of characterising the family of extreme value distributions is via the class of max-stable distributions.

Definition 1.3 (Max-stability) A non-degenerate distribution function $G$ is max-stable if $G^n$ is of the same type as $G$, for every $n = 2, 3, \ldots$.

In fact, $G$ is max-stable if and only if it is of the same type as an extreme value distribution.
Definition 1.4 (Domain of attraction) If there exist sequences of constants $a_n > 0$ and $b_n \in \mathbb{R}$, which satisfy (1.1), then $F$ is said to be in the domain of attraction of $G$.

Theorem 1.2 does not specify which (if any) of the three types will arise for a given $F$, nor does it identify which distributions $F$ are in the domain of attraction of each extreme value distribution. However, there exist necessary and sufficient conditions that resolve these issues (see for example Leadbetter et al., 1983 and Galambos, 1987).

1.3.3 The generalised extreme value distribution

For statistical purposes, it is inconvenient to work with three distinct classes of limiting distributions and it would be more useful to use a parameterisation which contains all three types. Von Mises (1954) and Jenkinson (1955) independently obtained such a parameterisation. Here, we present the generalised extreme value distribution (GEV).

Definition 1.5 (Generalised extreme value distribution) A random variable $X$ is said to follow a generalised extreme value distribution if

$$\Pr(X \leq x) = G(x) = \exp\left[-\{1 + \xi(x - \mu)/\sigma\}^{-1/\xi}\right],$$

where \(\{x\}_+ = \max(x, 0)\). The constants $\mu$, $\sigma(> 0)$ and $\xi$ denote the location, scale and shape parameters ($\xi$ is sometimes referred to as the extreme value index). When the above holds we write $X \sim GEV(\mu, \sigma, \xi)$.

The type 2 and type 3 classes of extreme value distribution correspond, respectively, to the cases $\xi > 0$ and $\xi < 0$ in this parameterisation, while the type 1 class arises in the limit when $\xi \to 0$. The relation between $\xi$ and the possible tail behaviours can be described in the following way: if $\xi < 0$ then the upper endpoint $w(G)$ is finite, whereas if $\xi = 0$ or $\xi > 0$ then $w(G)$ is infinite. Additionally, it is straightforward to show that the tail function $\overline{G}(x) = 1 - G(x)$ decreases exponentially when $\xi = 0$ and decreases polynomially when $\xi > 0$.

1.3.4 Point process characterisation

So far, the presented results have only concerned the limiting distribution of $M_n$ as $n \to \infty$. Statistical procedures based on these results are possible (see later) but
tend to be inefficient in that they are wasteful of the available extremal information. For example, the technique of modelling annual maxima with a generalised extreme value distribution is inefficient as a procedure for statistical inference because even though data may have been collected as often as daily or even hourly, only one observation per year is taken into account. In order to overcome this difficulty, several methods have been developed which include more of the extremal information than just the maximum $M_n$.

Firstly, techniques based on threshold methods were developed. These were then followed by methods based on the $r$-largest order statistics. Both of these approaches can be derived from the more general point process approach of Pickands (1971) which was developed further by Smith (1989). This latter result will be outlined in the following, but first we briefly recall some background results on point processes.

1.3.4.1 Point processes

A point process can be understood intuitively as a series of events occurring in time (or space) according to some statistical law. A point process $N$ on a rectangle $S$ in an $m$-dimensional Euclidean space can be defined as a family of non-negative integer (or $\infty$)-valued random variables, $N(B)$, indexed by the bounded Borel sets $B \subset S$. For a particular $B$, the random variable $N(B)$ denotes the number of events of $N$ that occur in $B$. See Leadbetter et al. (1983, pages 305-312).

Before stating the results, it is convenient to explore the concepts of point process convergence and recall the definition of a Poisson process.

Suppose that $\{N_n\}$ is a sequence of point processes on a rectangle $S$ (in an $m$-dimensional Euclidean space) and that $N$ is a point process. Then $N_n \xrightarrow{d} N$ when the sequence of random variables $\{N_n(B_1), \ldots, N_n(B_k)\}$ converges in distribution to $\{N(B_1), \ldots, N(B_k)\}$ for each choice of $k$, and all bounded Borel sets $B_i \subset S$ such that $N(\partial B_i) = 0$ almost surely, $i = 1, \ldots, k$ and where $\partial B$ represents the boundary of the set $B$.

A simple sufficient condition for point process convergence in distribution is given by Leadbetter et al. (1983, Theorem A.1, page 309). This result is a special case of a theorem of Kallenberg (1976).

A Poisson process $N$ is a point process, such that $N(B) \sim Po\{\Lambda(B)\}$, where $Po$ represents the Poisson distribution with intensity measure $\Lambda(B) = E(\text{number of points}$
in \( B \), where \( E \) denotes the expectation operator. Another characteristic of this process is that if \( B_1, \ldots, B_k \) are mutually disjoint Borel sets then \( N(B_i); \ i \leq k \) are independent Poisson random variables. If the point intensity \( \lambda \) (the derivative of \( \Lambda \)) is constant then \( N \) is an homogeneous Poisson process.

1.3.4.2 A point process result

Let \( X_1, \ldots, X_n \) be iid random variables with common distribution function \( F \), and suppose that \( F \) is in the domain of attraction of one of the extreme value distributions listed in Theorem 1.2. Consider the following sequence of 2-dimensional point processes:

\[
P_n = \left\{ \left( i/(n+1), (X_i - b_n)/a_n \right) : i = 1, \ldots, n \right\}.
\]

In these point processes non-extreme observations become scaled towards the lower boundary of the domain as \( n \to \infty \).

The process \( P_n \) converges weakly to a Poisson process \( P \) on sets which exclude the lower boundary. Thus, writing \( \Lambda \) for the integrated intensity measure of the limiting process, it follows from the Poisson property that for the region \( A = [0,1] \times (x, \infty) \),

\[
\exp \{-\Lambda(A)\} = \Pr(\text{no points in } A) = \Pr\left\{ \left( (M_n - b_n)/a_n < x \right) \right\} \approx \exp \left[ -\left\{ 1 + \xi(x - \mu)/\sigma \right\}^{1/\xi} \right].
\]

The last step utilises the usual statistical device of treating a probabilistic limit result as an approximation when \( n \) is sufficiently large. Here, the asymptotic limit \( G(x) \) is treated as the basis for an approximation of \( \Pr\left\{ (M_n - b_n)/a_n < x \right\} \). Therefore, at high levels, the process \( P_n \) should approximate a Poisson process with intensity measure on sets of the form \( \{(t_1, t_2) \times (x, \infty)\} \) given by

\[
\Lambda \left\{ (t_1, t_2) \times (x, \infty) \right\} = (t_2 - t_1) \left\{ 1 + \xi(x - \mu)/\sigma \right\}^{1/\xi},
\]

and with point intensity

\[
\lambda(t, x) = -\frac{\partial^2}{\partial t \partial x} \Lambda \left\{ (0, t) \times (x, \infty) \right\} = \sigma^{-1} \left\{ 1 + \xi(x - \mu)/\sigma \right\}^{-(1+\xi)/\xi}.
\]

Assuming that \( B = [b_1, x_1] \times \cdots \times [b_d, x_d] \subset \mathbb{R}^d \), then \( \Lambda(x_1, \ldots, x_d) = \frac{\sigma^d \Lambda(b_1)}{\sigma_1 \cdots \sigma_d} \).

This is easily shown by verification of the two conditions required by Kallenberg's result. Namely that there is convergence of both the expected number of points and the probability of no points in a certain region.
Essentially, this result gives a characterisation for all observations which are extreme in the sense of having exceeded a high threshold.

Taking \( u \) to be this high threshold, the likelihood of the point process for \( N_A \) observations on a region \( A = [0,1] \times (a, \infty) \) for \( x > u \), is given by

\[
L(A, \mu, \sigma, \xi) = \exp \{-\Lambda(A)\} \prod_{i=1}^{N_A} d\Lambda(t_i, x_i)
\]

\[
= \exp \left[-\{1 + \xi(x - \mu)/\sigma\}_+^{-1/\xi}\right] \prod_{i=1}^{N_A} \sigma^{-1}\{1 + \xi(x_i - \mu)/\sigma\}_+^{-(1+\xi)/\xi}
\]

where \( x_1, \ldots, x_{N_A} \) is an enumeration of the \( N_A \) points that exceed the threshold \( x \).

### 1.3.5 Threshold methods: the generalised Pareto distribution

The generalised Pareto distribution (GPD) may be derived via a simple application of the point process approach given above. The aim is to describe the limiting distribution of exceedances of a high threshold \( u \).

Let \( X_{i,n}^* = (X_i - b_n)/a_n \), and consider the conditional probability

\[
\Pr \left( X_{i,n}^* > x + u \mid X_{i,n}^* > u \right) = \Pr \left( \text{ith point of} \ P_n > x + u \mid \text{ith point of} \ P_n > u \right),
\]

where \( P_n \) denotes the point process given in (1.2).

Then, since \( P_n \xrightarrow{w} P \), this probability converges to the expected number of points of \( P \) exceeding \( x + u \) divided by the expected number exceeding \( u \). Thus,

\[
\Pr \left( X_{i,n}^* > x + u \mid X_{i,n}^* > u \right) \xrightarrow{n \to \infty} \frac{\Lambda \{[0,1] \times (x + u, \infty)\}}{\Lambda \{[0,1] \times (u, \infty)\}}.
\]

Finally, using equation (1.3), with \( \sigma' = \sigma + \xi(u - \mu) \) gives

\[
\lim_{n \to \infty} \Pr \left( X_{i,n}^* > x + u \mid X_{i,n}^* > u \right) = [1 + \xi x/\sigma']_+^{-1/\xi}.
\]

This is referred to as the generalised Pareto distribution with shape parameter \( \xi \in \mathbb{R} \) and scale parameter \( \sigma' > 0 \). This result was first established by Pickands (1975).

### 1.3.6 The \( r \)-largest order statistics method

This technique can also be derived as a special case of the point process approach.
Defining $M_n^{(i)}$ to be the $i$th largest order statistic, the goal here is to obtain the limiting joint distribution of

\[
\left( \frac{M_n^{(1)} - b_n}{a_n}, \frac{M_n^{(2)} - b_n}{a_n}, \ldots, \frac{M_n^{(r)} - b_n}{a_n} \right)
\]

for some high $r$. Note that $M_n^{(1)} = M_n$.

Setting $x = M_n^{(r)}$ in the likelihood (1.4) yields

\[
L(A, \mu, \sigma, \xi) = \exp \left[ - \left\{ 1 + \xi(M_n^{(r)} - \mu)/\sigma \right\}_+^{-(1+\xi)/\xi} \prod_{i=1}^{r} \sigma^{-1} \left\{ 1 + \xi(M_n^{(i)} - \mu)/\sigma \right\}_+^{-(1+\xi)/\xi} \right].
\]

Note that the region $A$ is such that $A = [0, 1] \times [M_n^{(r)}, \infty)$.

### 1.3.7 Stationary sequences

All the results derived so far have assumed independent and identically distributed series. However, since in several common areas of application observed data exhibit some form of temporal or spatial dependence, the consequence of dependence between the observed values of a sequence is an important issue. Thus, we now focus on stationary dependent sequences.

The importance of stationarity mainly concerns the following points. Firstly, satisfactory extremal theory can be developed, generalising the classical methods. Secondly, there is a wide range of applications for which stationary models are often more realistic than iid assumptions. Finally, even non-stationary sequences can often be divided into stationary periods which can be considered approximately independent.

**Definition 1.6 (Stationarity)** A sequence of random variables $X_1, X_2, \ldots$ is said to be stationary if the joint distributions of $(X_{j_1}, \ldots, X_{j_p})$ and $(X_{j_1+q}, \ldots, X_{j_p+q})$ are identical for every $p, j_1, \ldots, j_p$ and $q$.

In order to obtain useful results, conditions need to be imposed limiting the persistence of the dependence between $X_j$ and $X_{j+k}$ as $|k| \to \infty$. For this purpose two conditions $D$ and $D'$ were considered by Leadbetter et al. (1983). The condition $D$ restricts long range dependence, while the $D'$ condition restricts the dependence between neighbouring $X_j$'s (short-term dependence). Note that there are several different styles of dependence conditions in the literature (see e.g. O'Brien, 1987 or Leadbetter and Rootzén, 1998).
Definition 1.7 (The D(\(u_n\)) condition) For a given sequence \(u_n\), the condition \(D(u_n)\) is said to hold if for all \(i_1 < \cdots < i_p < j_1 < \cdots < j_q\) with \(j_q - i_p \geq l\)
\[
\frac{\Pr(X_{i_1} \leq u_n, \ldots, X_{i_p} \leq u_n, X_{j_1} \leq u_n, \ldots, X_{j_q} \leq u_n)}{\Pr(X_{i_1} \leq u_n, \ldots, X_{i_p} \leq u_n) \times \Pr(X_{j_1} \leq u_n, \ldots, X_{j_q} \leq u_n)} \leq \alpha_{n,l}
\]
where \(\alpha_{n,l} \to 0\) as \(n \to \infty\) for some sequence \(l_n = o(n)\).

In the following we let \(X_1, \ldots, X_n\) be a stationary sequence of random variables with common distribution function \(F\). For \(M_n\) defined as before, the following result holds:

Theorem 1.8 (Extremal types theorem for stationary sequences) If \(D(u_n)\) is satisfied with \(u_n = a_n x + b_n\) (with \(a_n > 0\) and \(b_n\) constants) and if
\[
\Pr\left\{\left(\frac{M_n - b_n}{a_n}\right) \leq x\right\} \xrightarrow{\mathcal{D}} G(x) \tag{1.7}
\]
for some non-degenerate distribution function \(G\), then \(G\) is a generalised extreme value distribution.

The \(D(u_n)\) condition requires that two separated groups of \(X\)'s have joint distribution function that factorises approximately into the joint distribution functions of each group as the separation \(l\) increases. In practice the level \(u_n\) is chosen to increase at a rate determined by the previous theorem, i.e. such that it stabilises (1.7).

The interpretation of Theorem 1.8 is that for stationary processes that are long-range independent (in the sense defined by the \(D(u_n)\) condition), the standard asymptotic limits apply for the maximum \(M_n\). That is, the dependence has no impact on the class of possible non-degenerate limit distributions. As for the iid case, this theorem does not inform which (if any) of the three types of limits apply for a given stationary process, even if \(F\) is known. Before addressing this issue, we require the following definitions:

Definition 1.9 (The D'(\(u_n\)) condition) The D'(\(u_n\)) condition is said to hold for a stationary sequence \(X_1, \ldots, X_n\) and sequence \(u_n\) of constants if
\[
\limsup_{n \to \infty} n \sum_{j=2}^{[n/k]} \Pr(X_1 > u_n, X_j > u_n) \to 0 \quad \text{as} \quad k \to \infty,
\]
where \([\cdot]\) denotes the integer part.
Alternatively, this condition may be expressed essentially in terms of conditional distributions which aids interpretation. It guarantees that extreme observations tend to occur singly at increasingly high levels, i.e. do not form clusters.

**Definition 1.10 (Associated independent sequence)** Let $X_1, \ldots, X_n$ be stationary random variables with common distribution function $F$, and let $X^*_1, \ldots, X^*_n$ be a sequence of iid random variables, each with distribution $F$. This sequence is called the independent sequence associated with $X_1, \ldots, X_n$ or the associated independent sequence.

The main results for (stationary) domains of attraction are as follows:

**Theorem 1.11** Suppose that $D(u_n)$ and $D'(u_n)$ are satisfied for the stationary sequence $X_1, \ldots, X_n$ where $u_n = a_n x + b_n$ for each $x$ ($a_n > 0$ and $b_n$ are the constants such that (1.7) holds) and define $M^*_n = \max(X^*_1, \ldots, X^*_n)$ for the associated independent sequence $X^*_1, \ldots, X^*_n$, then

$$\Pr \left\{ \left( \frac{M^*_n - b_n}{a_n} \right) \leq x \right\} \to G(x)$$

for some non-degenerate distribution function $G$ if and only if

$$\Pr \left\{ \left( \frac{M_n - b_n}{a_n} \right) \leq x \right\} \to G(x).$$

Thus, when the $D(u_n)$ and $D'(u_n)$ conditions are both satisfied, the normalising constants and the limit distributions are identical as for the iid case. It is important to note that this result says nothing about the rates of convergence.

Another important case to consider is the case when just $D(u_n)$ holds, i.e. when $D'(u_n)$ fails. This is the case of stationary sequences with long range extremal independence but local extremal dependence, so that exceedances of a high level can occur in clusters. First, we need to define the extremal index $\theta$.

**Definition 1.12 (The extremal index)** The sequence $X_1, \ldots, X_n$ has extremal index $\theta$, where $0 \leq \theta \leq 1$, if for each $\tau$,

1. There exists $u_n(\tau)$ such that $n \left[ 1 - F(u_n(\tau)) \right] \to \tau$, and

2. $\Pr \{ M_n \leq u_n(\tau) \} \to \exp(-\theta \tau)$. 

11
The extremal index is a measure of the amount of clustering in the process at extreme levels and can typically be interpreted as the reciprocal of the asymptotic mean size of extreme clusters. If a process is independent then the extremal index satisfies $\theta = 1$, but the converse is not true, i.e. $\theta = 1$ does not imply that the underlying observations are independent. In fact, any process which satisfies the $D'(u_n)$ condition has $\theta = 1$. The effect of the extremal index on the asymptotic behaviour of maxima is given by the following theorem.

**Theorem 1.13** Let $X_1, \ldots, X_n$ be a stationary sequence with extremal index $\theta > 0$ and consider $M^*_n$ as defined before. Then $M^*_n$ has a non-degenerate limiting distribution $G$ under linear normalisation, i.e.

$$\Pr \left\{ \left( M^*_n - b_n \right) / a_n \leq x \right\} \rightarrow G(x),$$

if and only if $M_n$ has limiting distribution $G^\theta$ with the same linear normalisation, i.e.

$$\Pr \left\{ \left( M_n - b_n \right) / a_n \leq x \right\} \rightarrow G^\theta(x) = G(a_\theta x + b_\theta).$$

Another interpretation of this result is that the limiting distributions are of the same type when the same normalising constants are used. Alternatively, but equivalently, if different normalising constants are used then the two limiting distributions may be taken as identical.

### 1.3.7.1 Point process results for stationary sequences

Consider $X_1, \ldots, X_n$ as being a stationary sequence satisfying slightly stronger versions of the $D(u_n)$ and $D'(u_n)$ conditions (see Leadbetter et al., 1983, sect. 5.4), and let $X^*_1, \ldots, X^*_n$ be the associated independent sequence. Then, the point process

$$P_n^* \equiv \left\{ \left\{ j/n, (X^*_j - b_n)/a_n \right\}; j = 1, \ldots, n \right\}$$

converges weakly to a non-homogeneous Poisson process with intensity measure given by equation (1.3) if and only if

$$P_n \equiv \left\{ \left\{ j/n, (X_j - b_n)/a_n \right\}; j = 1, \ldots, n \right\}$$

does too. That means that the limiting extremes from the sequence $X_1, \ldots, X_n$ behave like those from an iid sequence provided the process satisfies these long- and short-range extremal independence conditions.
Other point process results are available when the short-range condition does not hold, i.e. when there is local clustering of the extremes. Results concerning the point process of cluster occurrences and results on the number of exceedances and the associated exceedance point process can be found in Leadbetter (1999).

1.3.7.2 Non-stationary sequences

Extremal results for non-stationary and strongly dependent normal sequences have been developed. In particular, results have been obtained for non-stationary normal sequences with a diversity of possible mean and correlation structures (Leadbetter et al., 1983). Cases of processes consisting of a stationary normal sequence together with an added deterministic part, such as a seasonal component or trend, have been examined by Horowitz (1980), Hüsler (1981, 1986) and Leadbetter et al. (1983).

1.4 Multivariate extremes

We will now consider extremes of multivariate data. In contrast to the univariate case, where concepts such as maximum, order statistics and extreme values have natural definitions, in the multivariate case there are no natural definitions as several different concepts of ordering are reasonable and possible. The classical approach in this case has been to consider the maximum of the multivariate sample as being the vector of componentwise maxima.

1.4.1 Notation and limiting distribution

Suppose that \( X_1, \ldots, X_n \) are iid \( d \)-dimensional random variables with common joint distribution function \( F(x) = F(x_1, \ldots, x_d) \). Let the marginal distributions of \( F(x) \) be \( F_1, \ldots, F_d \) and define the \( d \)-vector \( M_n \) as the vector of componentwise maxima, i.e. \( M_n = (M_n^{(1)}, \ldots, M_n^{(d)}) \) where \( M_n^{(j)} = \max(X_n^{(j)}, \ldots, X_n^{(j)}) \) for \( j = 1, \ldots, d \) and \( X_i^{(j)} \) represents the \( j \)th component of \( X_i \).

As in the univariate case, the aim is to characterise the limiting behaviour of \( M_n \) or specifically the multivariate distribution \( G \) which may arise as a limit as \( n \to \infty \) in

\[
\Pr \left\{ \left( \frac{M_n^{(j)} - b_n^{(j)}}{a_n^{(j)}} \right) / x_j; j = 1, \ldots, d \right\} = F^n \left( u_n^{(1)}, \ldots, u_n^{(d)} \right) \to G(x) \quad (1.8)
\]
Chapter 1. Background

where \( u^{(j)}_n = a^{(j)}_n x_j + b^{(j)}_n \), for sequences \( a^{(j)}_n > 0 \) and \( b^{(j)}_n \in \mathbb{R}; 1 \leq j \leq d, n \geq 1 \).

The non-degenerate limiting distributions \( G \) are called multivariate extreme value distributions.

Equation (1.8) can be written in the simpler way

\[
\Pr \left\{ \left( M_n - b \right) / a \leq x \right\} = F_n (a_n x + b_n) \to G(x) \quad (1.9)
\]

using the obvious notation. Note that each marginal distribution \( G_j \) of the limiting joint distribution \( G \) is of extreme value type. Any multivariate distribution function \( F \) for which equation (1.9) holds is said to be in the domain of attraction of \( G \). See Resnick (1987) for details.

There are two separate limiting aspects in the limit in equation (1.9): the convergence of each marginal distribution of \( F^n \) to the corresponding marginal distributions of \( G \), and the convergence of the dependence structure of \( F^n \) to that of \( G \). These two aspects may be considered individually.

Issues relating to marginal convergence can be treated using the previous univariate results. Therefore, the dependence structure of \( G \), and the convergence of the dependence structure of \( F^n \) to that of \( G \) may be considered for standardised marginal variables. For simplicity of notation and clarity in presenting results, it is convenient to choose the marginal variables as unit Fréchet distributed, i.e. having distribution function \( \Pr(X \leq x) = \exp(-1/x) \) on \( x > 0 \). This does not result in a lack of generality, since univariate probability integral transformations can be used to relate the results to arbitrary marginal distributions.

1.4.2 A point process result for multivariate extremes

The point process result of de Haan (1985) presented in this section provides a simple characterisation of multivariate extreme value distributions.

As discussed above, \( X_1, \ldots, X_n \) are chosen to be iid \( d \)-dimensional random variables with unit Fréchet margins and distribution function \( F \), where \( F \) is in the domain of attraction of a multivariate extreme value distribution \( G \). The point process \( \mathcal{P}_n \) on \( \mathbb{R}_+^d \) is defined by

\[
\mathcal{P}_n = \left(n^{-1}X_i; i = 1, \ldots, n\right).
\]
The term $n^{-1}$ is the normalising constant needed to stabilise each componentwise maximum when the margins are unit Fréchet. Then, the point process $\mathcal{P}_n$ converges weakly to a non-homogeneous Poisson process $\mathcal{P}$ on $\mathbb{R}_+^d \setminus \{0\}$, in the limit as $n \to \infty$. Note that the origin is excluded from the domain of $\mathcal{P}$. This is because all the "non-extreme" observations $X_i$ become scaled to 0 in the limit so that an infinite number of points will concentrate there.

In order to obtain a simple expression for the point intensity of $\mathcal{P}$ it is convenient to use the pseudo radial and angular coordinates

$$ r_i = \frac{1}{n} \sum_{j=1}^{d} X_i^{(j)} \quad \text{and} \quad w_i^{(j)} = \frac{X_i^{(j)}}{nr_i}; \quad i = 1, \ldots, n; \quad j = 1, \ldots, d. $$

Then, the point intensity of $\mathcal{P}$ is

$$ \mu(dr \times dw) = r^{-2} dr \, dH(w) \tag{1.10} $$

where $H$ is a non-negative measure on the $(d-1)$-dimensional unit simplex

$$ S_d = \left\{ (w_1, \ldots, w_d); \sum_{j=1}^{d} w_j = 1, w_j \geq 0, \ j = 1, \ldots, d \right\} $$

satisfying

$$ \int_{S_d} w_j \, dH(w) = 1 \quad \text{for each} \ j = 1, \ldots, d. \tag{1.11} $$

This definition of the measure $H$ implies that no explicit parameterisation exists for such a measure.

By equation (1.10), the intensity $\mu$ factorises into a known function of the radial component and a measure of the angular component. This shows clearly that the radial and angular components ($r$ and $w$) of the process $\mathcal{P}$ are independent.

Note that the previous point process result provides a characterisation which includes dependence between extremes of the marginal variables using the dependence measure $H$.

### 1.4.3 Multivariate extreme value distribution

We now present a representation of a general multivariate extreme value distribution with unit Fréchet margins that is a consequence of the point process result given above.
Consider \( A = \mathbb{R}_+^d \setminus \{(0, x_1) \times \cdots \times (0, x_d)\} \). Then, since \( \mathcal{P}_n \overset{w}{\to} \mathcal{P} \) as \( n \to \infty \), where \( \mathcal{P} \) is a Poisson process, we have that

\[
\Pr(n^{-1} \mathbf{X}_i \notin A, i = 1, \ldots, n) \to \exp \{-\mu(A)\}
\]

where
\[
\mu(A) = \int_A r^{-2} \, dr \, dH(w) = \int_{S_d} \int_0^\infty \min_{1 \leq i \leq d} \left( \frac{s_i}{x_i} \right) r^{-2} \, dr \, dH(w)
\]
\[
= \int_{S_d} \max_{1 \leq j \leq d} \left( \frac{w_j}{x_j} \right) \, dH(w).
\]

Considering the equality\(^4\)

\[
\Pr(n^{-1} \mathbf{X}_i \notin A, i = 1, \ldots, n) = \Pr\left(n^{-1} \mathbf{M}_n^{(j)} \leq x_j, j = 1, \ldots, d\right),
\]

and combining with equations (1.9) and (1.13), it follows that any limit distribution of normalised componentwise maxima with unit Fréchet margins is of the form \( G(x) = \exp \{-V(x)\} \) where

\[
V(x) = V(x_1, \ldots, x_d) = \int_{S_d} \max_{1 \leq j \leq d} \left( \frac{w_j}{x_j} \right) \, dH(w)
\]

for some measure \( H \) as defined above. This is Pickands's (1981) representation theorem for multivariate extreme value distributions with unit Fréchet margins. The function \( V \) defined in equation (1.14) is called the dependence function.

Since \( V \) is determined by the measure \( H \) which has no closed form representation, there exists no explicit parameterisation for \( V \), and consequently, no closed form representation of a general multivariate extreme value distribution.

It is easy to see from equation (1.14) that \( V \) is a homogeneous function of order \(-1\) (i.e. \( V(nx) = n^{-1}V(x) \)) and so \( G(nx) = G(x) \), i.e. if \( G \) exists, then \( G \) is in its own domain of attraction and \( G \) is max-stable. A necessary and sufficient condition for the multivariate distribution \( F \) to be in the (iid) domain of attraction of a multivariate extreme value distribution \( G \) is given by Resnick (1987, Props 5.15 and 5.17).

Coles and Tawn (1991) related the dependence function \( V \) and the measure density \( h \) in the following way

\[
\frac{\partial^d V}{\partial x_1 \cdots \partial x_d} = -\left( \sum_{i=1}^d x_i \right)^{-(d+1)} h\left( \frac{x_1}{\sum x_i}, \ldots, \frac{x_d}{\sum x_i} \right).
\]
Chapter 1. Background

Note that the measure density $h$ in equation (1.15) is the corresponding density of $H$ on the interior of the unit simplex $S_d$. The corresponding densities on each of the lower dimensional boundaries of $S_d$ can be also related to the dependence function $V$ in a similar way, see Coles and Tawn (1991).

1.4.4 Examples

So far attention has been given to the general structure and properties of multivariate extreme value distributions. Since there exists no general parameterisation for such distributions we will instead use some illustrative examples and will focus on the bivariate case for simplicity.

For clarity, as discussed before, the marginal variables are chosen to be Fréchet distributed. Then, the distribution is determined by its dependence structure, i.e. by the measure $H$ in equation (1.10) or equivalently the dependence function $V$ in equation (1.14).

In the bivariate case, for a given measure $H$ satisfying the normalisation condition (1.11), the corresponding dependence function $V$ is then given by

$$V(x, y) = \int_0^1 \max \left( \frac{w}{x}, \frac{1-w}{y} \right) dH(w)$$

$$= \frac{1}{y} \int_{\frac{x}{x+y}}^1 (1-w) \, dH(w) + \frac{1}{x} \int_0^{\frac{x}{x+y}} w \, dH(w).$$

Some parametric examples of such functions $V$ and $H$ now follow:

1. The logistic model

The dependence function is

$$V(x, y) = \left( x^{-1/\alpha} + y^{-1/\alpha} \right)^{\alpha}, \quad (1.16)$$

and the measure density is given by

$$h(w) = (\alpha^{-1} - 1) \left( w^{-1/\alpha} + (1-w)^{-1/\alpha} \right)^{\alpha-2} \left( w (1-w) \right)^{-(1+1/\alpha)},$$

for $w \in [0, 1]$, where the dependence parameter $\alpha \in (0, 1]$. The cases $\alpha = 1$ and $\alpha \to 0$ correspond to exact independence and complete dependence respectively. The logistic model has been widely used in statistical modelling in the
bivariate case due to its simple and tractable form. This model is one of the main
dependence structures used in this thesis.

The multivariate logistic model can be found in Gumbel (1960).

2. The mixed model

The measure \( H \) is given by

\[
H(w) = \begin{cases} 
1 - \theta & \text{if } w = 0 \\
1 - \theta + 2\theta w & \text{if } 0 < w < 1 \\
2 & \text{if } w = 1
\end{cases}
\]

so that the measure has equal mass \( 1 - \theta \) at positions \( w = 0 \) and \( w = 1 \). The
_corresponding dependence function is

\[
V(x, y) = \frac{1}{x} + \frac{1}{y} - \frac{\theta}{x + y},
\]

where \( 0 \leq \theta \leq 1 \). Independence corresponds to \( \theta = 0 \), but complete
dependence cannot be represented.

For a multivariate version of the mixed model see page 48.

3. The asymmetric logistic model

This model, proposed by Tawn (1990), has dependence function

\[
V(x, y) = \left( (\frac{x}{\theta})^{-1/\alpha} + \left( \frac{y}{\phi} \right)^{-1/\alpha} \right) \alpha + \frac{1 - \theta}{x} + \frac{1 - \phi}{y},
\]

and measure \( H \) with density given by

\[
h(w) = \frac{1 - \alpha}{\alpha \theta \phi} \left\{ \left( \frac{w}{\theta} \right)^{-1/\alpha} + \left( \frac{1 - w}{\phi} \right)^{-1/\alpha} \right\} \alpha \left\{ \frac{w}{\theta} \left( \frac{1 - w}{\phi} \right) \right\}^{-(1+1/\alpha)},
\]

for \( 0 < w < 1 \) and with atoms of mass \( 1 - \phi \) and \( 1 - \theta \) at positions \( w = 0 \)
and \( w = 1 \), respectively, where \( \theta, \phi \in [0, 1] \) and \( \alpha \in (0, 1] \).

The multivariate version of the asymmetric logistic model can be found in

Other examples can be found in Coles and Tawn (1991, 1994) and Joe (1989).
1.5 Statistical modelling

We consider here the implementation of the previous theoretical extreme value results in statistical modelling.

When using extreme value statistical techniques for data analysis, the goal is to model the upper tail of $F$. A fundamental assumption is that the underlying distribution $F$ is in the domain of attraction of an extreme value distribution, and moreover, that the limiting distribution is representative, and may be used as a model, of the underlying process in some tail region.

1.5.1 Fitting the univariate models

Fitting statistical models requires the estimation of parameters or other model features using the information conveyed by the observed data. The two main schools of methodology within extreme value statistical modelling are parametric approaches, where a parameter is estimated by, for example, maximum likelihood or the method of moments, and non-parametric approaches, which do not depend on a specific model.

For applied statistical methods, likelihood based methods are often used and we will concentrate on them. The theory of likelihood based estimation is well understood, and inferences are easily modified to incorporate more complex model structures. There is a technical complication in using maximum likelihood to fit univariate extreme value models, since the regularity conditions of maximum likelihood are satisfied only when the shape parameter $\xi > -1/2$. This issue was studied by Smith (1985), and alternative estimation schemes were presented for other values of $\xi$. The situation $\xi < -1/2$, which corresponds to distributions with a very short (and bounded) upper tail, is not very common with environmental data, thus, in practice, maximum likelihood estimation is valid and regular as a procedure for inference.

1.5.2 Univariate statistical methods

There are two main approaches for modelling univariate extreme values; the peaks over threshold method based on the generalised Pareto distribution of Pickands (1975) and a method based on the point process approach of Pickands (1971).
Consider \( X_1, \ldots, X_n \) as being iid univariate variables with common unknown distribution function \( F \) and let \( x_1, \ldots, x_n \) represent the observed univariate data sequence. Let \( u \) denote a high threshold in the sense that observations which exceed \( u \) are extreme and those which do not exceed \( u \) are non-extreme. The aim here is to model the form of \( F(x) \) on \( x \geq u \).

When using likelihood methods to fit a model the usual approach is based on censoring. We consider that an extreme point, i.e. a point that exceeds the threshold \( u \), directly contributes information about the upper tail of the process, while non-extreme points do not provide information about the form of the distribution above \( u \). The relevant extremal information conveyed by the points below \( u \) is the relative frequency of extreme and non-extreme points. Thus, it is natural to consider the non-extreme points as censored at the threshold \( u \).

Note that the threshold \( u \) must be chosen carefully. Indeed, a threshold that is too low may lead to biased estimation due to the invalidity of the asymptotic argument, whereas a threshold that is too high will have few exceedances and lead to low precision results. In practice, it is important to check that there is reasonable stability in results and conclusions for a range of different choices of the threshold \( u \).

### 1.5.2.1 The peaks over threshold (POT) method

In this case, exceedances of the high threshold \( u \) are assumed to occur following a generalised Pareto distribution as in equation (1.6). Those that are below \( u \) are considered censored at \( u \). As a result, the model for the upper tail of the distribution \( F \) is given by

\[
\Pr(X \leq x) = 1 - \lambda_u \left\{ 1 + \frac{\xi(x-u)}{\sigma'} \right\}^{-1/\xi} \quad \text{for} \quad x \geq u \tag{1.17}
\]

where \( \lambda_u = 1 - F(u) \), \( \sigma' > 0 \) is a scale parameter and \( \xi \) is a shape parameter.

Each point \( x_i > u \) makes a contribution to the likelihood \( L(\lambda_u, \sigma', \xi; x_i) \) given by

\[
L(\lambda_u, \sigma', \xi; x_i) = \frac{dF(x)}{dx} \bigg|_{x=x_i} = \lambda_u \sigma'^{-1} \left\{ 1 + \frac{\xi(x_i-u)}{\sigma'} \right\}^{-1-1/\xi}\left(1+\xi\right)/\xi.
\]

Due to the censoring scheme, the likelihood contribution of a point \( x_i \) which fails to exceed the threshold \( u \) is given by

\[
L(\lambda_u, \sigma', \xi; x_i) = F(u) = 1 - \lambda_u.
\]
Therefore, the overall likelihood for a set of $n$ iid points is

$$L_n(\lambda_u, \sigma', \xi; x) = \prod_{i=1}^{n} L(\lambda_u, \sigma', \xi; x_i)$$

$$= \left\{ \prod_{i:x_i \leq u} (1 - \lambda_u) \right\} \left\{ \prod_{j:x_j > u} \lambda_u \sigma'^{-1} \{1 + \xi (x_j - u)/\sigma'\}^{-(1+\xi)/\xi} \right\}$$

$$= (1 - \lambda_u)^{n - n_u} \lambda_u^{n_u} \sigma'^{-n_u} \prod_{j:x_j > u} \{1 + \xi (x_j - u)/\sigma'\}^{-(1+\xi)/\xi}$$

where $n_u$ is the number of points that exceed the threshold $u$.

### 1.5.2.2 Point process approach

As detailed in Section 1.3.4.2, realisations of $X$ which exceed the threshold $u$ are assumed to follow a non-homogeneous Poisson process with intensity measure for the set $(\alpha, \infty)$ given by

$$\Lambda(x) = \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi} \quad \text{for} \quad x \geq u,$$

where $\mu$ and $\sigma > 0$ are location and scale parameters respectively, and $\xi$ is the same shape parameter as in equation (1.17).

Using again the fact that non-extreme values are considered censored at the threshold, the likelihood contribution of a point $x_i$ is given by

$$L(\mu, \sigma, \xi; x_i) = \begin{cases} \exp \{-\Lambda(u)\} & \text{if } x_i \leq u, \\ \exp \{-\Lambda(u)\} \lambda(x_i) & \text{if } x_i > u, \end{cases}$$

where $\lambda(x) = \Lambda'(x)$. The overall likelihood for the $n$ iid observations is therefore

$$L_n(\mu, \sigma, \xi; x) = \exp \left\{ -n \{1 + \xi(x - \mu)/\sigma\}^{-1/\xi} \right\} \prod_{j:x_j > u} \sigma^{-1} \{1 + \xi (x_j - \mu)/\sigma\}^{-(1+\xi)/\xi}.$$

The $r$-largest order statistics model can also be fitted in a similar way, with the threshold chosen to be the $r$th largest order statistic.

### 1.5.2.3 Distribution-free methods

In this section, the extreme type limits considered are of the form:

$$F_n(a_n x + b_n) \rightarrow \exp \left\{ -(1 + \xi x)^{-1/\xi} \right\} \quad \text{as } n \rightarrow \infty.$$
The location and scale parameters usually contained in the generalised extreme value representation are absorbed into the normalising constants $a_n$ and $b_n$. We present here some of the distribution-free estimators available for the shape parameter $\xi$ and the normalising constants $a_n$ and $b_n$. These estimators are based on the top $k$ order statistics.

1. Hill’s estimator

This estimator for the shape parameter $\xi$, which is valid only for $\xi > 0$, is defined by

$$\hat{\xi}^+ = k^{-1} \sum_{i=0}^{k-1} (\log X_{n-i:n} - \log X_{n-k:n})$$

where $X_{n-i:n}$ is the $i$th largest order statistic. See Hill (1975).

2. de Haan’s estimator

This is a refinement of the Hill estimator which is valid for a general $\xi$. Defining

$$R_n(r) = k^{-1} \sum_{i=0}^{k-1} (\log X_{n-i:n} - \log X_{n-k:n})^r \quad \text{for } r = 1, 2,$$

the estimator is defined by

$$\hat{\xi} = R_n(1) + \frac{1}{2} \left\{ \frac{1 - (R_n(1) / R_n(2))^2}{R_n(1)} - 1 \right\}^{-1}.$$  

Note that $R_n(1)$ is the Hill estimator given above. See de Haan (1994).

3. Estimators for $a_n$ and $b_n$

Estimators for the normalising constants $a_n$ and $b_n$ are given by

$$\hat{a}_n = X_{n-k:n}R_n^{(1)} \max \left(1, 1 - \hat{\xi} \right) \quad \text{and} \quad \hat{b}_n = X_{n-k:n},$$

where $R_n^{(1)}$ and $\hat{\xi}$ are as defined above. Alternatively, see Dekkers et al. (1989).

1.5.3 Multivariate statistical methods

As previously mentioned, the marginal and dependence structure aspects of multivariate extreme value results may be considered individually. This is also true for the
associated statistical methods. In this section, dependence structure modelling will be focused on.

For simplicity, only the case of iid multivariate observations is considered and, as discussed in Section 1.4.1, marginal variables are assumed to be unit Fréchet distributed, since a simple marginal transformation can be used when this is not the case. In practice, the tails of the marginal distributions are modelled using standard univariate extreme methods within the multivariate estimation.

Methods for modelling the dependence structure are outlined next. All assume that asymptotic results may be taken as exact for some suitable region of the joint tail.

For clarity, only the bivariate case will be considered. Let $F$ be the common distribution function of the iid bivariate variables $X_1, \ldots, X_n$, where $F$ is in the domain of attraction of the bivariate extreme value distribution $G$.

### 1.5.3.1 Multivariate point process approach

In applying the de Haan (1985) result given in Section 1.4.2, realisations of $X$ which occur in a region of the joint tail that is sufficiently far from the origin are assumed to follow a non-homogeneous Poisson process $P$ with point intensity given by

$$
\mu(dr \times dw) = r^{-2} dr dH(w).
$$

This approach therefore requires modelling the unknown measure $H$. This can be done adopting a parametric family for $H$, such as the logistic model in Section 1.4.4.

### 1.5.3.2 Multivariate threshold models

Resnick (1987) gives a result stating that the $d$-dimensional distribution function $F$ is in the domain of attraction of a multivariate extreme value distribution function $G$ with unit Fréchet margins if and only if

$$
\lim_{t \to \infty} \frac{1 - F(tx_1, \ldots, tx_d)}{1 - F(t, \ldots, t)} = \frac{-\log G(x_1, \ldots, x_d)}{-\log G(1, \ldots, 1)} = \frac{V(x_1, \ldots, x_d)}{V(1, \ldots, 1)}.
$$

A threshold based model for multivariate extreme values is developed in Smith et al. (1997) by assuming that the limit in equation (1.20) holds exactly for large values of $t$, or equivalently, for a fixed $t$ and large values of $(x_1, \ldots, x_d)$. Thus, this model
applies when each marginal variable exceed a high threshold. Ledford and Tawn (1996) developed a multivariate extreme value model, which incorporates marginal independence as a special case, based on an equivalent assumption. In the same work, in order to include asymptotic independence as a special case, they obtained a new model for bivariate extreme tails. In this model, the asymptotic form of the joint survivor function $F$ of an arbitrary random pair $(X, Y)$, with unit Fréchet margins, is modelled directly by

\[ \Pr(X > r, Y > r) = F(r, r) = C(r)^{-1/\eta} \]  

for large $r$ (1.21)

where $C(r)$ is a slowly varying function (i.e. $C(tr)/C(r) \to 1$ as $r \to \infty$ for all fixed $t > 0$) and $\eta$, the coefficient of tail dependence, lies in the range $(0, 1]$.

Later, Ledford and Tawn (1997) proposed a flexible joint asymptotic expansion of $F(r_1, r_2)$ that contained and extended model (1.21). Considering $r_1$ and $r_2$ simultaneously large, they modelled the joint survivor function $F(r_1, r_2)$ by

\[ F(r_1, r_2) = C_1(r_1, r_2)r_1^{-c_1}r_2^{-c_2} + C_2(r_1, r_2)r_1^{-(c_1+d_1)}r_2^{-(c_2+d_2)} + \ldots \]  

(1.22)

where $c_1 + c_2 = 1/\eta$, $d_j \geq 0$ and $C_j(r_1, r_2) \neq 0$ denote bivariate slowly varying functions.

Conditions under which the components of this expansion are uniquely determined are also specified in Ledford and Tawn (1997). Modelling the bivariate slowly varying function $C_1(r_1, r_2)$ is also considered in this paper.

This characterisation provides a smooth transition between perfect dependence and exact independence and allows negative dependence between the marginal extremes. The flexibility of this framework is very important for statistical modelling, which will be studied in Chapters 4 and 5.

\subsection{Fitting the multivariate models}

As in the univariate case, there are two schools of methodology within multivariate extreme value statistical modelling: parametric and non-parametric.

The major advantage of parametric modelling is that standard likelihood techniques permit simultaneous fitting of marginal and dependence models producing improved parameter estimates and measures of precision. In contrast, non-parametric approaches require that each margin is modelled separately and then the dependence model is estimated. This approach does not allow exchange of information between the margins,
nor does it take into account the uncertainty of the marginal estimation when the dependence structure is estimated.

**Multivariate censored likelihood**

To illustrate this method we use the bivariate case. Let \( \{(x_i, y_i); i = 1, \ldots, n\} \) be a sequence of independent observations of a bivariate random variable \((X, Y)\). Suppose that the distribution function of \((X, Y)\) is known to be of the form \( F_\theta(x, y) \), where \( \theta \) is a vector of unknown parameters, for observations which exceed a pair of high thresholds \((u_1, u_2)\).

To estimate \( \theta \) from the observed data, a bivariate version of the previously used censoring scheme is implemented since the parametric model is valid only when \( X > u_1 \) and \( Y > u_2 \).

It is convenient to divide the outcome space into the four regions given by

\[
\{R_{kl}; \ k = I(X > u_1), l = I(Y > u_2)\},
\]

where \( I \) is the indicator function. Thus, for a marginal observation which does not exceed the threshold, the only relevant information it conveys is that it occurs below the threshold and not its actual value (Smith et al., 1997 and Ledford and Tawn, 1996).

The likelihood contribution, \( L_{kl} \), for a point \((x_i, y_i)\) which falls in region \( R_{kl} \) is then given by

\[
L_{00}(x_i, y_i) = F_\theta(u_1, u_2), \quad L_{01}(x_i, y_i) = \frac{\partial F_\theta(u_1, y_i)}{\partial y} \bigg|_{y = y_i},
\]

\[
L_{10}(x_i, y_i) = \frac{\partial F_\theta(x_i, u_2)}{\partial x} \bigg|_{x = x_i}, \quad L_{11}(x_i, y_i) = \frac{\partial^2 F_\theta(x_i, y_i)}{\partial x \partial y} \bigg|_{(x, y) = (x_i, y_i)}.
\]

The likelihood for a set of \( n \) independent points is therefore

\[
L_n(\theta) = \prod_{i=1}^{n} \left[ \sum_{k,l \in \{0,1\}} L_{kl}(x_i, y_i) I \{(x_i, y_i) \in R_{kl}\} \right],
\]

which may be maximised using standard numerical methods.

This method can be extended easily to \( d \)-dimensions; the likelihood contribution for a typical observation \((y_1, \ldots, y_d)\) in which only the components \( j_1, \ldots, j_m \) exceed their thresholds is given by

\[
\frac{\partial^m F_\theta(x_1, \ldots, x_d)}{\partial x_{j_1} \cdots \partial x_{j_m}} \bigg|_{\{x_j = \max(u_j, y_j); j = 1, \ldots, d\}}.
\]
Again, estimation proceeds by standard numerical methods.

**Non-parametric approach**

Consider again that $X$ and $Y$ are unit Fréchet distributed and that $F$ is in the domain of attraction of a bivariate extreme value distribution $G$ with unit Fréchet margins. Then for large values of $x$ and $y$,

$$F(x, y) \approx G(x, y) = \exp \{-V(x, y)\} \approx 1 - V(x, y) \quad (1.24)$$

or, since $V$ is homogeneous of order $-1$, then

$$1 - F(tx, ty) \approx V(tx, ty) = t^{-1}V(x, y) \approx t^{-1}\{1 - F(x, y)\}.$$  

It follows immediately that if $A \subseteq B$, where $B = \mathbb{R}^2_+ \setminus \{ (0, x) \times (0, y) \}$, then

$$\Pr \{(X, Y) \in tA\} \approx t^{-1} \Pr \{(X, Y) \in A\} \quad (1.25)$$

for all $t \geq 1$.

In practical applications, equation (1.25) is taken to hold exactly for large values of $x$ and $y$ and suitable values of $t$. This homogeneity property, which is also valid for arbitrary tail events, provides a means of relating the probability of events more extreme than any observed to that of extreme events within the observed data. The basis of such methods is therefore to use equation (1.25) to scale estimates of probabilities within a set $A$ that contains data so as to obtain probability estimates for the set $tA$ which contains no data. For further details see Huang (1992) and Draisma and de Haan (1995).

This procedure may bring practical difficulties when there is asymptotic independence between the extreme marginal variables and for certain shapes of the "failure region" $A$, e.g. the region $R_{11}$ (see Coles et al., 1999). These difficulties may be avoided by extending the model in (1.24) to take into account the degree of dependence within the class of asymptotically independent distributions. The Ledford and Tawn (1997) model in equation (1.22) may then be used to approach this problem. In this case, the analogous result to (1.25) is given by

$$\Pr \{(X, Y) \in tA\} \approx t^{-1/\eta} \Pr \{(X, Y) \in A\}$$

for $t \geq 1$. Apart from the additional issue of the estimation of $\eta$, considerations for inference remain the same as for the previous procedure (Coles et al., 1999).
Inference for the limiting point process model described in Section 1.4.2 involves estimating the measure $H$ in equation (1.10) by an empirical estimator which uses points that are sufficiently far from the origin (de Haan, 1985). Alternative non-parametric procedures have been proposed by Einmahl et al. (1997) and de Haan and de Ronde (1998).

1.6 Dependence measures for extreme values

The focus of this section is the development of measures of extremal dependence for bivariate random variables $(X, Y)$.

1.6.1 The measures $\chi$ and $\overline{\chi}$

Assuming that the marginal distribution of $X$ and $Y$ are identical, one natural measure is

$$\chi = \lim_{z \to z^*} \Pr(Y > z \mid X > z), \quad (1.26)$$

where $z^*$ is the upper limit of the support of the common marginal distribution. This measure was considered as far back as Sibuya (1960) and Tiago de Oliveira (1962/63). In the case $\chi = 0$, the componentwise maxima of iid copies of $(X, Y)$ are said to be asymptotically independent, and are said to be asymptotically dependent otherwise (i.e. $\chi > 0$) whenever the limit in (1.26) exists. Less formally, the marginal variables are described as asymptotically independent or asymptotically dependent in these cases.

This measure is generalised by Coles et al. (1999) to the case of non-identically distributed pairs $(X, Y)$ by transformation to Uniform margins $(U_1, U_2)$. In their work, the copula function is used in order to remove the influence of the marginal aspects when measuring dependence. Let $F$ be the distribution function of the pair $(X, Y)$ and $F_1$ and $F_2$ the respective marginal distribution functions. The copula function $C(\cdot, \cdot)$ is such that

$$F(x, y) = C \{F_1(x), F_2(y)\},$$

with domain $D = [0, 1] \times [0, 1]$. Thus, $C$ is the joint distribution after transformation to Uniform $[0,1]$ margins via $(U_1, U_2) = \{F_1(X), F_2(Y)\}$. Further details can be seen in Nelsen (1998) and Joe (1997).
The measure $\chi$ then has the form

$$\chi = \lim_{t \to 1} \Pr(U_2 > t \mid U_1 > t),$$

or, equivalently,

$$\chi = \lim_{t \to 1} \left\{ 2 - \frac{1 - \Pr(U_1 < t, U_2 < t)}{1 - \Pr(U_1 < t)} \right\} = \lim_{t \to 1} \left\{ 2 - \frac{\log \Pr(U_1 < t, U_2 < t)}{\log \Pr(U_1 < t)} \right\} = \lim_{t \to 1} \chi(t)$$

with $0 \leq \chi \leq 1$. The function $\chi(t)$ can also be interpreted as a quantile-dependent measure. For instance, its sign determines whether the variables are positively or negatively associated at the quantile level $t$ (Coles et al., 1999).

The class of asymptotically independent distributions has notorious importance in multivariate extreme value modelling. However, since $\chi = 0$ for all members of this class, no information on relative strength of dependence is given for such models by this measure. To overcome this limitation a new dependence measure $\bar{\chi}$ is defined in Coles et al. (1999) by

$$\bar{\chi} = \lim_{t \to 1} \left\{ \frac{2 \log \Pr(U_1 > t)}{\log \Pr(U_1 > t, U_2 > t)} - 1 \right\} = \lim_{t \to 1} \left\{ \frac{2 \log(1-t)}{\log \bar{C}(t,t)} - 1 \right\}$$

where $\bar{C}$ is such that $\bar{C}(F_1(x), F_2(y)) = F(x, y)$.

The measure $\bar{\chi}$ falls within the range $[-1, 1]$, with the set $[-1, 1)$ corresponding to asymptotic independence and the point 1 to asymptotic dependence.

In conclusion, as a summary of extreme dependence, the complete pair $(\chi, \bar{\chi})$ is required. The case $(\chi > 0, \bar{\chi} = 1)$ represents asymptotic dependence, where the value of $\chi$ determines a measure of the strength of dependence within this class. The case $(\chi = 0, \bar{\chi} < 1)$ represents asymptotic independence, where the value of $\bar{\chi}$ now determines the strength of dependence within this class.

### 1.6.2 The coefficient of tail dependence $\eta$ and $\mathcal{L}(r)$

Ledford and Tawn (1996) show that under broad conditions the joint survivor function of an arbitrary random pair $(X, Y)$ with unit Fréchet margins satisfies the asymptotic condition (1.21), or equivalently

$$\Pr(X > r, Y > r) \sim \mathcal{L}(r) \{\Pr(X > r)\}^{1/\eta} \text{ for large } r,$$  \hspace{1cm} (1.27)
where $L(r)$ is a slowly varying function and $\eta$, the coefficient of tail dependence, lies in the range $(0, 1]$.

The bounding cases of perfect negative and positive dependence correspond respectively to $\eta \to 0$ and $\eta = 1$ with $L(r) = 1$. For exactly independent variables, $\eta = 1/2$ and $L(r) \sim 1$. The componentwise maxima are asymptotically dependent if $\eta = 1$ with $L(r) \to c > 0$ as $r \to \infty$ and asymptotically independent if $\eta < 1$. Also, if $1/2 < \eta \leq 1$ the marginal variables are positively associated and if $0 < \eta < 1/2$ the marginal variables are negatively associated. Hence, the parameter $\eta$ characterises the nature of the tail dependence, and $L$ its relative strength for a given $\eta$.

### 1.6.3 Relationship between $\eta$ and $L$ and the dependence measures $\chi$ and $\overline{X}$

We will now examine how the dependence measures $\chi$ and $\overline{X}$ may be derived from the coefficient of tail dependence $\eta$ and $L$.

It follows from (1.27) and the definition of $\overline{C}$ that

$$\overline{C}(t,t) \sim L \left\{ (1 - t)^{-1} \right\} (1 - t)^{1/\eta} \text{ as } t \to 1.$$ 

Hence,

$$\overline{X} = \lim_{t \to 1} \left\{ \frac{2 \log(1 - t)}{\log L \left\{ (1 - t)^{-1} \right\} + \frac{1}{\eta} \log(1 - t)} - 1 \right\} = 2\eta - 1.$$

The relationship between these pairs of measures is then given by

$$\overline{X} = 2\eta - 1$$

$$\chi = \begin{cases} c \text{ if } L(t) \to c \ (0 < c \leq 1) \text{ and } \overline{X} = 1 \\ 0 \text{ if } \overline{X} < 1 \text{ or } (\overline{X} = 1 \text{ and } L(t) \to 0) \end{cases}.$$

Estimators for $\eta$ and $L(t)$ and hence for $\chi$ and $\overline{X}$ may be found in Leadford and Tawn (1996) and Coles et al. (1999).

### 1.7 Examples of application areas

Extreme value theory has found widespread applications in many different areas. Primarily it was used in modelling environmental extremal phenomena and typically those
which may substantially damage property and impact on people's lives. Examples include the analysis of wind speeds (see for example Anderson and Turkman, 1992), sea-levels (see for example Coles and Tawn, 1991) and river heights (Gumbel and Mustafi, 1967). This theory has also been used as a tool in studying air pollution problems. An example of this application is given by Smith (1989) where a point process based modelling approach is used in the study of ozone.

Extreme value techniques can be used in the modelling of spatial extremes. Casson and Coles (2000), developed a model for the space-time evolution of hurricane wind-fields and addressed the issue of spatial dependence in extremes of hurricane events.

Applications have also been made in finance and insurance. Some examples are the study of extreme stock market price movements by Longin (1996) where the asymptotic distribution of extreme returns was estimated, and the analysis of the size-distribution of yearly claims in insurance by Zajdenweber (1994).

Another important body of extreme value theory research is the characterisation of the extremal behaviour of Markov chains. This theory provides a major advantage in the applicability of extreme value techniques to time series data (see for example Bortot and Coles, 1999).

In the field of reliability and survival analysis, Goka (1994) studied the distribution of the smallest and/or largest values of the tensile strength and the life of a certain material. Other studies, such as that of Facchini and Spinelli (1994), concern the problem of the collapse risk of a rigid block under a seismic excitation. More engineering applications are discussed extensively in the book by Castillo (1987), which includes several sets of data.
Chapter 2

Assessment and comparison of the various estimation procedures

In univariate statistical modelling the upper tail of an unknown distribution function is modelled from given observed data using a range of statistical models, e.g. the GEV distribution, the peaks over threshold (POT) model, the r-largest model or the point process model, as detailed in Chapter 1. Whichever of these models are used, at some point they need to be fitted to the observed data, and various estimation methods are available for doing this. In this chapter, we examine the performance of these various estimation methods in practice and comment on their performance. In the following we examine fitting the peaks over threshold (POT) model using the parametric likelihood method, the semiparametric methods of Hill and de Haan, the probability weighted moments (pwm) method and the Bayesian approach of MCMC (with a flat prior).

2.1 Introduction

When fitting the POT model, that is, assuming that exceedances of a high threshold \( u \) occur following a generalised Pareto distribution, two parameters need to be estimated: the scale parameter \( \sigma \) and the shape parameter \( \xi \). Different estimation methodologies described in the literature will be compared here. We will start with the parametric approach, using first the maximum likelihood method, as in Section 1.5.2.1, where standard errors and covariances associated with the parameter estimates are calculated using the observed Fisher information (Smith, 1984). The other parametric approach
Chapter 2. Assessment and comparison of the various estimation procedures

we examine is the method of probability weighted moments (PWM) of Hosking and Wallis (1987). This method consists of equating $\alpha_r$, the theoretical $r$th probability weighted moment of a variable $X$ assumed having a generalised Pareto distribution function $F$, to $\hat{\alpha}_r$, the empirical probability weighted moment based on a sample of size $n$. That means, equating

$$\alpha_r = E[X\{1 - F(X)\}^r] = \frac{\sigma}{(r + 1)(r + 1 - \xi)}, \quad \text{for } r = 0, 1, \quad (2.1)$$

to

$$\hat{\alpha}_r = n^{-1} \sum_{j=1}^{n} x_{jm\{1 - \hat{F}(x_{jm})\}^r}, \quad r = 0, 1,$$

where $\hat{F}$ is an empirical estimate of the distribution function $F$, and $x_{1:n} \leq \ldots \leq x_{n:n}$ is the ordered sample. Note that the moment $\alpha_r$ defined in (2.1) exists only when $\xi < 1$. Here, we use $\hat{F}(x_{jm}) = (j - 0.35)/n$ which is the empirical estimate of the distribution function recommended by Hosking and Wallis (1987). Asymptotic formulae for the standard errors and covariances associated with the parameter estimates are also given in that paper. We remark that, in order to use this method, exceedances have to be shifted so that the threshold of the resulting data set (of exceedances) is zero.

Next, we consider a non-parametric estimation approach: the Hill and the de Haan estimators for the shape parameter $\xi$ (defined in Section 1.5.2.3, equations (1.18) and (1.19), respectively). We use the Hill estimator when $\xi > 0$ and the de Haan estimator otherwise, since this latter estimator is valid for all $\xi \in \mathbb{R}$. In both cases, the estimator for the scale parameter $\sigma$ is given by,

$$\hat{\sigma} = \frac{X_{n-kn}R_n^{(1)}}{\rho_1(\xi)}, \quad (2.2)$$

where $X_{n-kn}$ is the $k$th largest order statistic, $R_n^{(1)}$ represents the Hill estimator for the shape parameter and

$$\rho_1(\xi) = \begin{cases} 1 & \text{for } \xi \geq 0 \\ \frac{1}{1-\xi} & \text{for } \xi < 0 \end{cases}$$

(see de Haan, 1994). Standard errors and covariances for these estimators can be derived from the representation of the limit random variables as linear combinations of normals, see de Haan and Rootzén (1993).

The last approach used to fit the POT model to data is the Markov chain Monte Carlo (MCMC) method. Here we use the Metropolis-Hastings algorithm (see for example
Chapter 2. Assessment and comparison of the various estimation procedures

Tanner, 1996) with a flat prior and proposals based on a bivariate Normal random walk \(^1\) to estimate the scale and shape parameters. We remark that any reasonable flat prior should have a minor effect on our results. Estimates are calculated using the posterior mean of the MCMC output.

All these methods are implemented using S-Plus, apart from the MCMC method which is undertaken using Fortran. To illustrate these methods, data will be simulated from the unit Fréchet, normal and exponential distributions. This choice was made in order to work with data sets whose normalised maximum distributions converge to a generalised extreme value distribution at different rates. In each case, we work with samples of size 2500 and 5000, and results are obtained for two different thresholds taken at the threshold probabilities 0.9 and 0.95. On average, this results in 250 and 125 threshold exceedances when the sample size is 2500, and 500 and 250 when the sample size is 5000.

Note that the estimation of extreme quantiles is often the main requirement of an extreme value analysis, and therefore, this is often the key factor for judging the performance of different estimation methods. Motivated by this, we consider estimates of the \(1 - p\) quantile, \(q_p\), where

\[
q_p = u + \frac{\sigma}{\xi} \left( \left( \frac{\lambda u}{\beta} \right)^{-\xi} - 1 \right),
\]

obtained by inverting equation (1.17). In extreme value terminology, \(q_p\) is the return level associated with the return period \(1/p\), and it is common to extrapolate equation (2.3) to return levels substantially beyond the range of the data to which the model has been fitted.

2.2 Analysis of simulated data

2.2.1 Simulated data

The simulated data sets and respective thresholds used in this analysis are the following:

- Two unit Fréchet data sets with sample size equal to 2500 and 5000, respectively.

\(^1\)The bivariate normal distribution is centred at the current parameter values and has covariance matrix chosen by ad hoc means.
The two different thresholds $u = -1/\log p$, where the threshold probability $p$ takes the values 0.9 and 0.95, are used.

- Two exponentially distributed data sets with sample size 2500 and 5000. The thresholds here are $u = -\log(1 - p)$, where $p$ takes the values 0.9 and 0.95.
- Two normal data sets with sample size equal to 2500 and 5000, respectively. The thresholds used are $u = \Phi^{-1}(p)$ for $p = 0.9$ and $p = 0.95$, where $\Phi$ is the standard normal distribution function.

It is well known that the Extremal types theorem (Theorem 1.2, Section 1.3.2) holds for each of these distributions. In fact, the normalised maximum of iid unit Fréchet variables has an exact unit Fréchet distribution for all $n \geq 1$. A Gumbel distribution is the limit of the normalised maximum of iid normal variables although convergence is very slow, see Leadbetter et al. (1983), page 39. For the exponential, the Gumbel again results as the limiting distribution with a convergence rate that is intermediate to the Fréchet and Normal. Using these results and the relationship between the GEV parameters and those from the generalised Pareto distribution, as in Section 1.3.5, it is easy to derive the true values of the POT model parameters, $\xi$ and $\sigma$, for the Fréchet and exponential distributions. Sub-asymptotic values of these parameters for the normal case can be derived from the unpublished work “Approximation in extreme value theory” of Smith. The limiting values of these parameters are shown in the following table for each distribution.

<table>
<thead>
<tr>
<th>Fréchet</th>
<th>Exponential</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi$</td>
<td>$1$</td>
<td>$0$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>$u$</td>
<td>$1$</td>
</tr>
</tbody>
</table>

Table 2.1: Limiting values of the POT shape and scale parameters for the Fréchet, exponential and normal distributions. Here $\phi$ is the standard normal density function.

### 2.2.2 Results and analysis

For the Fréchet case, the Hill estimator is chosen to be used over the de Haan estimator since it is known that $\xi > 0$. However, had we used the de Haan estimator then similar results would have been obtained. With the exponential and normal data examples, the
Chapter 2. Assessment and comparison of the various estimation procedures

Table 2.2: Results for Fréchet data: Shape and scale parameter estimates with associated standard errors (given in parentheses) obtained by the maximum likelihood method, the Hill estimator and MCMC. Recall that for the MCMC method estimates are posterior means using a uniform prior.

de Haan estimator is used, since $\xi = 0$. In Tables 2.2, 2.3 and 2.4, parameter estimates and corresponding standard errors obtained for the different methods are shown for each of the underlying distributions.

Results for pwm are not given in Table 2.2 since, in this Fréchet case, the true value of the shape parameter $\xi$ is 1 and $\alpha_r$ in equation (2.1) does not exist and therefore the method is not strictly appropriate. However, if this difficulty is ignored, then the parameter estimates obtained by this method are good but there are no values for the associated standard errors since the formulae given in Hosking and Wallis (1987) break down and yield negative variances.

In general, the performance of these methods seems to be good, mainly in the exponential and Fréchet cases, since parameter estimates have values close to the true values. The associated standard errors are also small, apart from those obtained for Fréchet data which are particular large for the scale parameter estimates. In the Fréchet case, the Hill estimator seems to be favoured since it gives estimates for the shape parameter closer to one and small standard errors. In the exponential case, the maximum likelihood method seems to perform well, whereas in the normal case the MCMC method seems to perform best. Interestingly, parameter estimates and their associated standard errors using the maximum likelihood and MCMC methods are very similar. More-
Chapter 2. Assessment and comparison of the various estimation procedures

Table 2.3: Results for exponential data: Shape and scale parameter estimates with associated standard errors (given in parentheses) obtained by the maximum likelihood method, the pwm method, the de Haan estimator and MCMC.

<table>
<thead>
<tr>
<th>n</th>
<th>p</th>
<th>u</th>
<th>MLE</th>
<th>pwm</th>
<th>de Haan</th>
<th>MCMC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(\hat{\xi} )</td>
<td>(\hat{\sigma} )</td>
<td>(\hat{\xi} )</td>
<td>(\hat{\sigma} )</td>
</tr>
<tr>
<td>2500</td>
<td>0.9</td>
<td>2.30</td>
<td>0.09</td>
<td>0.97</td>
<td>0.11</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.07)</td>
<td>(0.09)</td>
<td>(0.07)</td>
<td>(0.09)</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>3.00</td>
<td>0.06</td>
<td>1.00</td>
<td>0.10</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.11)</td>
<td>(0.15)</td>
<td>(0.10)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>5000</td>
<td>0.9</td>
<td>2.30</td>
<td>-0.08</td>
<td>1.09</td>
<td>-0.02</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.05)</td>
<td>(0.07)</td>
<td>(0.05)</td>
<td>(0.07)</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>3.00</td>
<td>-0.05</td>
<td>1.07</td>
<td>-0.06</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.05)</td>
<td>(0.11)</td>
<td>(0.08)</td>
<td>(0.13)</td>
</tr>
</tbody>
</table>

Table 2.4: Results for normal data: Shape and scale parameter estimates with associated standard errors (given in parentheses) obtained by the maximum likelihood method, the pwm method, the de Haan estimator and MCMC. Table 2.1 gives that when \(u = 1.28\), 'true' \(\sigma = 0.57\), and when \(u = 1.64\), 'true' \(\sigma = 0.48\).

<table>
<thead>
<tr>
<th>n</th>
<th>p</th>
<th>u</th>
<th>MLE</th>
<th>pwm</th>
<th>de Haan</th>
<th>MCMC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(\hat{\xi} )</td>
<td>(\hat{\sigma} )</td>
<td>(\hat{\xi} )</td>
<td>(\hat{\sigma} )</td>
</tr>
<tr>
<td>2500</td>
<td>0.9</td>
<td>1.28</td>
<td>-0.14</td>
<td>0.55</td>
<td>-0.11</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.07)</td>
<td>(0.05)</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>1.64</td>
<td>-0.17</td>
<td>0.53</td>
<td>-0.16</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.08)</td>
<td>(0.06)</td>
<td>(0.11)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>5000</td>
<td>0.9</td>
<td>1.28</td>
<td>-0.12</td>
<td>0.51</td>
<td>-0.18</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.05)</td>
<td>(0.04)</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>1.64</td>
<td>-0.09</td>
<td>0.44</td>
<td>-0.14</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.05)</td>
<td>(0.03)</td>
<td>(0.08)</td>
<td>(0.04)</td>
</tr>
</tbody>
</table>
over, results obtained by these two methods are generally better than those from the

pwm method. Shape parameter estimates using the de Haan estimator are also good

but the same is not true when using the scale parameter estimator (2.2). In fact, for

exponential and normal data, scale parameter estimates obtained by this estimator are

poorer and the associated standard errors are also higher. Similar conclusions are ob­

tained when using the Hill estimator, although scale parameter estimates obtained by

equation (2.2) are not so poor for the Fréchet case. As expected, we see that estimates

improve when larger samples are used and that estimates are worse when using very

high thresholds due to the reduced number of exceedances used to fit the model. Com­

paring the values in Tables 2.2, 2.3 and 2.4, we can conclude that parameter estimates

are closer to the true values and standard errors are lower when using exponential data.

The poorest results seems to be obtained when using normal data, where it can be noti­

ced that shape parameter estimates are all negative. That negative shape parameter

estimates are obtained in this case is to be expected, since it is known (see for example

Fisher and Tippett, 1928) that, in practice, it is more sensible to use a Weibull dis­

tribution (instead of a Gumbel) as an approximation of the maximum distribution for

the Gaussian and a sequence of shape parameters \( \xi_n < 0 \) such that \( \xi_n \uparrow 0 \) as \( n \to \infty \).

A graphical diagnostic to check the quality of the fit using these different methods

is given in Figures 2.1, 2.2 and 2.3 for the Fréchet, exponential and normal cases,

respectively. In these figures, only the case where the sample size is 2500 and threshold

probability \( p = 0.9 \) is presented.

Both the probability plot and the quantile plot suggest that the quality of the model

is good for almost all the calibration methods for the three data sets. However, the

quantile plot is more sensitive to slight departures from the model accuracy in the upper
tail than the probability plot. Apart from those obtained by the de Haan methodology,
results obtained using the remaining methods are very similar and suggest a good

modelling quality.

The return level plot, which also includes 95% confidence intervals, shows how the

fitted model extrapolates from the sample information. For the maximum likelihood

method, the confidence interval is based on the profile likelihood since this procedure
allows the interval to be asymmetric, a property which the delta-method does not
capture adequately. For the Hill, de Haan and the pwm cases, the confidence intervals
are based on the delta-method, whereas for the MCMC case, they are obtained using
Figure 2.1: Results for Fréchet data: Diagnostic plots of the model fit using a) the maximum likelihood method, b) the Hill estimator and c) the MCMC method. Note that results are presented in logarithmic scale and therefore $f_1$ (in the density plot) represents the density function of the logarithm of the exceedances. The asymmetric confidence interval in the return level plot of case b) is due to this logarithmic scale. The sample has size 2500 and the threshold probability used is $p = 0.9$. 

38
Figure 2.2: Results for exponential data: Diagnostic plots of the model fit using a) the maximum likelihood method, b) the pwm method, c) the de Haan estimator and d) the MCMC method. The sample has size 2500 and the threshold probability used is \( p = 0.9 \).
Chapter 2. Assessment and comparison of the various estimation procedures

Figure 2.3: Results for normal data: Diagnostic plots of the model fit using a) the maximum likelihood method, b) the pwm method, c) the de Haan estimator and d) the MCMC method. The sample has size 2500 and the threshold probability used is $p = 0.9$. 

40
the usual Monte Carlo approach for inference, i.e., they are constructed directly from
the sample \((q_p)_i = g(\xi_i, \sigma_i)\), where \(g\) is the function defined in equation (2.3). Once
more, extrapolation of the return levels suggests that the fit is reasonably good with the
exception of results using de Haan’s method. However, confidence intervals in the return
level plot obtained when using the Hill, de Haan or the pwm methodology appear to be
too wide. The negative values of the shape parameter estimates for the normal case are
reflected by the concave extrapolation in the return level plots. In the exponential case,
since the shape parameter estimates are very close to zero, extrapolation on the return
level plots is near-linear. Positiveness of shape parameter estimates in the Fréchet case
is also reflected in the convexity of the return level plots (note that the axes in this
case are on a log scale).

The final figure compares the histogram of the exceedances data with the fitted gen­
eralised Pareto density function. Results improve when using samples with size 5000
and are slightly better when using the threshold probability \(p = 0.9\). Again, results
obtained for the exponential data appear slightly better.

The adequacy of the fitted model as well as the chosen threshold can be checked using
plots of the ensemble of parameter estimates obtained through varying the threshold.
When fitting the model over a range of thresholds, we should observe some stabil­
ity in the parameter estimates (relative to their sampling variability) for thresholds
where the asymptotic arguments are valid. However, since the scale parameters in the
Fréchet and normal cases are threshold dependent we needed to do some corrections
to their estimates in order to get stability (see Coles, 1999). Plots of parameter esti­
mates against threshold, together with 95% confidence bands based on their standard
errors are shown in Figures 2.4, 2.5 and 2.6. Estimates are calculated using the differ­
et calibration methods for samples with size 2500 and 5000. The true values of the
parameters are represented by a line.

From the analysis of the three figures we again notice that when using a bigger sam­
ple the estimates are closer to the true values and that the standard errors increase
for higher thresholds and decrease for larger sample sizes. Parameters estimates and
standard errors are very similar for both the likelihood and MCMC methods for each
data set. Apart from the scale parameter estimates given by the estimator (2.2) re­
sults are quite good for the exponential case, and it appears that this is actually the
case where results are best. As noticed before, the scale parameter estimates using the

41
Figure 2.4: Results for Fréchet data: Scale and shape parameter estimates against threshold in the POT model analysis with sample sizes 2500 (above) and 5000 (below) together with 95% confidence bands. The different methods used are a) maximum likelihood, b) Hill estimator and c) MCMC.
Figure 2.5: Results for exponential data: Scale and shape parameter estimates against threshold in the POT model analysis with sample sizes 2500 (above) and 5000 (below) together with 95% confidence bands. The different methods used are a) maximum likelihood, b) pwm, c) de Haan estimator and d) MCMC.
Figure 2.6: Results for normal data: Scale and shape parameter estimates against threshold in the POT model analysis with sample sizes 2500 (above) and 5000 (below) together with 95% confidence bands. The different methods used are a) maximum likelihood, b) pwm, c) de Haan estimator and d) MCMC.
estimator (2.2) are not very good for exponential and normal data but seem better for Fréchet data. In the normal case, we observe that shape parameter estimates are always negative and considerably far from the limiting value zero. In this case, scale parameter estimates are also poor.

2.3 Summary and conclusions

The quality of the fit of the POT model is generally good. The calibration methods with better results are maximum likelihood and MCMC, and for these methods the results are very similar. The methodology based on the de Haan estimator produces poor estimates for the scale parameter that tend to be smaller than the true value. This results in the poor quality of the fitted model. As expected, parameter estimates improve when larger samples are used and estimates are worse when using very high thresholds due to the reduced number of exceedances. Similarly, the standard errors decrease when larger samples are used and increase with higher thresholds.

The quality of the fit is better when using exponential data since parameter estimates are closer to the true values and associated standard errors are small in this case. Parameter estimates using normal data are poorer and have larger standard errors. However, in spite of this, the quality of the model fit to the observed exceedances is quite good. The adequacy of the results obtained by all the calibration methods, except those obtained by the methodology based on the de Haan estimator, is confirmed by the model extrapolation. Again, the difficulties encountered when extrapolating the model obtained for the de Haan's case are due to the bad results obtained for the scale parameter estimates.
Chapter 3

Score tests of independence in multivariate extreme values

All existing score tests of independence in multivariate extreme values have non-regular properties that arise due to violations of the usual regularity conditions of maximum likelihood. Some of these violations may be dealt with using standard techniques, for example when independence corresponds to a boundary point of the parameter space of the underlying model. However, another type of regularity violation, the infinite second moment of the score function, is more difficult to deal with and has important consequences for applications, resulting in score statistics with non-standard normalisation and poor rates of convergence. The corresponding tests are difficult to use in practical situations because their asymptotic properties are unrepresentative of their behaviour for the sample sizes typical of applications, and extensive simulations may be needed in order to evaluate their null distributions in such cases. Overcoming this difficulty is the primary focus of this chapter.

We propose a likelihood based approach that provides asymptotically normal score tests of independence with regular normalisation and rapid convergence. The resulting tests are straightforward to implement and are beneficial in practical situations with realistic amounts of data.
Chapter 3. Score tests of independence in multivariate extreme values

3.1 Introduction

When studying multivariate extreme value distributions two separate aspects have to be considered: the component marginal distributions, which can be understood using univariate extreme value methods, and the dependence structure that relates them. Our focus here will be the dependence structure and, specifically, developing score tests to identify the important special case when the marginal variables are independent. Other techniques for testing independence may be found in the literature, e.g. a Cramer-von Mises-type statistic by Deheuvels and Martynov (1996), a dependence function based test by Deheuvels (1980), a test based on the number of points below certain thresholds by Dorea and Miasaki (1993) and the Kendall's tau test (see e.g. Capéra et al., 1997). However none of them seems to have been used in the literature.

Score tests of independence in multivariate extreme value distributions have been developed by Tawn (1988) and Ledford and Tawn (1996). Their results yield non-regular likelihood estimation frameworks that provide, via a non-regular normalisation, score statistics with asymptotically normal null distributions. However, close inspection of these tests reveals them to have poor rates of convergence, a problem which leads to difficulties when they are used in practice. For example, extensive simulation may be required in order to obtain good estimates of critical points of the test statistic for each particular sample size of interest. In this chapter we overcome this and associated difficulties by constructing asymptotically normal score tests of independence with regular normalisation and rapid convergence.

3.2 Score tests of independence

As we are primarily interested in dependence issues here, we will follow the framework of Ledford and Tawn (1996) and work with standardised marginal variables that we choose, for simplicity, to be unit Fréchet distributed. As mentioned before, this approach admits no loss of generality, as probability integral transformations may be used in order to extend this framework to arbitrary marginal distributions.

Let $X_1, \ldots, X_n$ be iid $d$-dimensional random variables with joint distribution function $F$ and unit Fréchet marginal distributions, and consider the $d$-vector of componentwise maxima $M_n$, as defined in Chapter 1. We assume that $F$ is a multivariate extreme
value (MEV) distribution so that

\[ \Pr (M_n/n < x) = F(x) \quad \text{for all } n \geq 1 \]

with \( F(x) = \exp \{ -V(x) \} \), where \( V \) is the dependence function defined in (1.14). We will concentrate here on the two most widely used dependence models, the logistic and the mixed models, that is, with dependence functions defined by

\[ V(x) = \left( \sum_{i=1}^{d} x_i^{-1/\alpha} \right)^\alpha \]

for the dependence parameter \( \alpha \in (0, 1] \) and

\[ V(x) = \sum_{i=1}^{d} x_i^{-1} + (-1)^{d+1} \theta \left( \sum_{i=1}^{d} x_i \right)^{-1} \]

for \( \theta \in [0, 1] \), respectively. Note that independence of the marginal variables arises when \( \alpha = 1 \) and when \( \theta = 0 \) for the logistic and mixed models respectively, which are both boundary points of the respective parameter spaces.

Non-regular score tests based on these specified parameter values were developed by Tawn (1988) and Ledford and Tawn (1996). The modelling framework that underpins these tests is the starting point of our development of regular score tests. For simplicity of presentation, we restrict attention to the bivariate case, \( d = 2 \), as it is clear how to obtain corresponding modifications in the multivariate \( (d \geq 3) \) case, and restricting attention to the bivariate case admits no loss of generality when testing for independence in multivariate extreme value distributions (Tiago de Oliveira, 1962/63). We will focus on the logistic dependence structure for \( V(x) \) in our derivation, and then state some corresponding results for the mixed model.

Let \((X_1, Y_1), \ldots, (X_n, Y_n)\) denote iid bivariate random variables with joint distribution function \( F \) and unit Fréchet marginal distributions. It is convenient to partition the outcome space \( R = \{(x, y) : (x, y) \in \mathbb{R}^+ \times \mathbb{R}^+\} \) into the four regions given by

\[ \{ R_{kl} : k = I(x \geq u), l = I(y \geq u) \} \quad (3.1) \]

where \( u \geq 0 \) is a threshold and \( I \) denotes the indicator function. The different modelling frameworks adopted by Tawn (1988) and Ledford and Tawn (1996) may be described using the following common representation: the joint distribution function \( F \) is assumed to satisfy

\[ F(x, y) = \exp \left\{ - \left( x^{-1/\alpha} + y^{-1/\alpha} \right)^\alpha \right\} \quad (3.2) \]

for \( \alpha \in (0, 1] \) and \((x, y)\) in some specified region \( R_\ast \), say.

The Tawn (1988) approach takes \( R_\ast = R \), and thus treats the distribution function (3.2) as a model for the entire outcome space of \((X, Y)\). In contrast, the approach adopted
Chapter 3. Score tests of independence in multivariate extreme values

by Ledford and Tawn (1996) takes $R_* = R_{11}$ as in definition (3.1) where $u$ is chosen to be a high quantile of the unit Fréchet distribution, and so assumes that the distribution function (3.2) holds only within a joint tail region of $R$. In order to obtain score tests of independence under both approaches we need to consider the corresponding likelihood functions. We will denote the likelihoods under the Tawn (1988) and Ledford and Tawn (1996) approaches by $L_n^{(1)}$ and $L_n^{(2)}$ respectively.

Under the Tawn (1988) approach, since the distribution function (3.2) is assumed to hold over the whole of $R$, the likelihood contribution of each observation is the joint density $f(x, y) = \partial^2 F(x, y)/\partial x \partial y$, and thus $L_n^{(1)}(\alpha) = \prod_{i=1}^n f(X_i, Y_i)$. The corresponding score function for $\alpha = 1$ therefore is given by

$$U_n^{(1)} = \left. \frac{\partial}{\partial \alpha} \log L_n^{(1)}(\alpha) \right|_{\alpha=1} = \sum_{i=1}^n S_0^{(1)}(X_i, Y_i),$$

where

$$S_0^{(1)}(X_i, Y_i) = (1 - X_i^{-1}) \log X_i + (1 - Y_i^{-1}) \log Y_i + (2 - X_i^{-1} - Y_i^{-1}) \log (X_i^{-1} + Y_i^{-1}) - (X_i^{-1} + Y_i^{-1})^{-1}. \quad (3.3)$$

Under the Ledford and Tawn (1996) approach, the distribution function (3.2) is treated as a model for only joint exceedances of a threshold $u$, that is for $R_* = R_{11}$. In developing their model, Ledford and Tawn (1996) considered the marginal observations below the threshold $u$ as censored at the threshold. Thus, the likelihood contribution, $L_n^{(2)}$, corresponding to a point $(x, y)$ which falls in region $R_{kl}$, is given by

$$L_n^{(2)}(x, y) = F(u, u) \quad L_n^{(2)}(x, y) = \frac{\partial F}{\partial x}(x, u)$$

$$L_n^{(2)}(x, y) = \frac{\partial F}{\partial y}(u, y) \quad L_n^{(2)}(x, y) = f(x, y), \quad (3.4)$$

and therefore, the likelihood for a set of $n$ independent points is given by

$$L_n^{(2)}(\alpha) = \prod_{i=1}^n \left( \sum_{k, l \in \{0, 1\}} I \{ (X_i, Y_i) \in R_{kl} \} L_{kl}^{(2)}(X_i, Y_i) \right).$$

Proceeding as above, the corresponding score function for $\alpha = 1$ is given by $U_n^{(2)} = \sum_{i=1}^n \sum_{k, l \in \{0, 1\}} I \{ (X_i, Y_i) \in R_{kl} \} S_{kl}^{(2)}(X_i, Y_i)$ where

$$S_{00}^{(2)}(X_i, Y_i) = -2 u^{-2} \log 2$$

$$S_{01}^{(2)}(X_i, Y_i) = -u^{-1} \log u + (1 - Y_i^{-1}) \log Y_i + (1 - u^{-1} - Y_i^{-1}) \log(u^{-1} + Y_i^{-1})$$

$$S_{10}^{(2)}(X_i, Y_i) = -u^{-1} \log u + (1 - X_i^{-1}) \log X_i + (1 - X_i^{-1} - u^{-1}) \log(X_i^{-1} + u^{-1})$$

$$S_{11}^{(2)}(X_i, Y_i) = (1 - X_i^{-1}) \log X_i + (1 - Y_i^{-1}) \log Y_i + (2 - X_i^{-1} - Y_i^{-1}) \log(X_i^{-1} + Y_i^{-1}) - (X_i^{-1} + Y_i^{-1})^{-1}. \quad (3.5)$$
Note that the likelihood contributions obtained under both the Tawn (1988) and Ledford and Tawn (1996) approaches coincide for observations in region $R_{11}$, and the same is true for the corresponding score contributions.

When $\alpha = 1$, i.e. when the marginal variables are independent, it can be shown that both $U_n^{(1)}$ and $U_n^{(2)}$ have expectation zero, consistent with regular likelihood theory, and infinite variance, which is inconsistent with regular likelihood theory. This infinite variance produces non-regular behaviour in the score statistics based on $U_n^{(1)}$ and $U_n^{(2)}$, as shown by the following.

**Proposition 3.1** (Tawn, 1988; Ledford and Tawn, 1996) Let $c_n = \{(n \log n)/2\}^{\frac{1}{2}}$. If the marginal variables are independent, then, as $n \to \infty$, both of the following hold:

$$-U_n^{(1)}/c_n \overset{w}{\to} N(0,1) \quad \text{and} \quad -U_n^{(2)}/c_n \overset{w}{\to} N(0,1),$$

where $\overset{w}{\to}$ denotes convergence in distribution.

The minus signs in the above test statistics are present so that positive dependence yields positive values of the test statistics. The normalisation here, $c_n = \{(n \log n)/2\}^{\frac{1}{2}}$, is heavier than that of regular cases (which have normalisation $n^{\frac{1}{2}}$) and provides the extra scaling necessary to counter the infinite variance of $U_n^{(1)}$ and $U_n^{(2)}$ in order to obtain a normal limit. It is important to note that these asymptotic results say nothing about the rate of convergence of $c_n^{-1}U_n^{(1)}$ and $c_n^{-1}U_n^{(2)}$ to standard normal distributions. For the above tests to be most useful in practice, we would hope for fast convergence, but simulation shows that this is not the case and in fact convergence to the standard normal distribution is very slow in both cases. Thus, in order to apply these tests in practical situations extensive simulations may be required in order to evaluate the appropriate critical points of their null distributions for each sample size of interest.

### 3.2.1 Regularised score tests of independence

For practical purposes, implementation of the test statistics given in Proposition 3.1 can be problematic because of their slow convergence to standard normal distributions, a property that arises because of the infinite variances of $U_n^{(1)}$ and $U_n^{(2)}$. Careful analysis shows that it is the presence of the $(x^{-1} + y^{-1})^{-1}$ term in the score function for a region that extends to $(\infty, \infty)$ that leads to the infinite variance in both cases. This, in
turn, may be attributed to the joint density \( f(x, y) \) being the likelihood contribution of observations within such a region. Our proposal for obtaining score tests with regular normalisation is to change this feature of the Tawn (1988) and Ledford and Tawn (1996) frameworks by censoring region \( R_{11} \) under both approaches, so that the only information exploited within the resulting likelihoods about the observations in \( R_{11} \) is how many of them there are. With this censoring in place we obviously lose information, but this is done as a compromise to obtain tests that have much faster convergence to standard normal distributions and are thus more applicable in practice. A contrasting technique would be to modify the test statistics by discarding the \( (x^{-1} + y^{-1})^{-1} \) term, an approach that has been investigated by Kimber and Zhu (1999) and Kimber and Sarker (2002), however this results in a test statistic with non-zero expectation.

Applying this censoring procedure to the Tawn (1988) framework we obtain the following likelihood

\[
L_n^{(1*)}(\alpha) = \prod_{i=1}^{n} \left[ I\{ (X_i, Y_i) \notin R_{11} \} f(X_i, Y_i) + I\{ (X_i, Y_i) \in R_{11} \} \bar{F}(u, u) \right] \tag{3.6}
\]

where \( \bar{F}(u, u) = \Pr(X > u, Y > u) = 1 - 2 \exp(-u^{-1}) + \exp(-2u^{-1}) \). The corresponding score function for \( \alpha = 1 \) is given by

\[
U_n^{(1*)} = \sum_{i: (X_i, Y_i) \notin R_{11}} S_0^{(1)}(X_i, Y_i) + \frac{2u^{-1} \log 2 \exp(-2u^{-1}) N_{11}}{2 \exp(-u^{-1}) - \exp(-2u^{-1}) - 1}
\]

for \( S_0^{(1)}(X_i, Y_i) \) as in equation (3.3) and where \( N_{11} \) denotes how many of the \( n \) observations fall in region \( R_{11} \). Similarly, censoring \( R_{11} \) under the Ledford and Tawn (1996) framework, we obtain the likelihood

\[
L_n^{(2*)}(\alpha) = \prod_{i=1}^{n} \left[ \sum_{i: (X_i, Y_i) \notin R_{11}} I\{ (X_i, Y_i) \in R_{kl} \} L_k^{(2)}(X_i, Y_i) + N_{11} \bar{F}(u, u) \right] \tag{3.7}
\]

and hence the score statistic

\[
U_n^{(2*)} = \sum_{i: (X_i, Y_i) \notin R_{11}} I\{ (X_i, Y_i) \in R_{kl} \} S_k^{(2)}(X_i, Y_i) + \frac{2u^{-1} \log 2 \exp(-2u^{-1}) N_{11}}{2 \exp(-u^{-1}) - \exp(-2u^{-1}) - 1}
\]

for \( I_k^{(2)} \) and \( S_k^{(2)} \) as defined in equations (3.4) and (3.5), respectively. The score functions \( U_n^{(1*)} \) and \( U_n^{(2*)} \) have zero expectation, as before, but more importantly for our purposes have finite variances, which we denote by \( n \sigma_1^2 \) and \( n \sigma_2^2 \) respectively. More precisely, by the central limit theorem, we have the following:
Chapter 3. Score tests of independence in multivariate extreme values

**Proposition 3.2** If the variables are independent, then, as \( n \to \infty \), both of the following hold:

\[
-U_n^{(1*)}/\sqrt{n} \xrightarrow{m} N(0,1) \quad \text{and} \quad -U_n^{(2*)}/\sqrt{n} \xrightarrow{m} N(0,1),
\]

where \( \sigma_1^2 \) and \( \sigma_2^2 \) denote the variances of the corresponding modified score statistics of a single point.

Thus we have obtained tests of independence with regular normalisation. In Figure 3.1 the convergence rates of these test statistics are examined informally and, more importantly, compared to those of the existing Tawn (1988) and Ledford and Tawn (1996) approaches. Clearly, the test statistics corresponding to the unmodified approaches have empirical distributions that are some considerable way from their limiting common \( N(0,1) \) law, whereas those corresponding to the modified approaches are much closer to the \( N(0,1) \) distribution. The results here are for the single sample size \( n = 300 \), but additional simulations, not reported here, show that these findings remain true even for much larger sample sizes. This suggests that the tests obtained via our \( R_{11} \) censoring modification have rapid convergence to \( N(0,1) \), and hence that tests based on \( U_n^{(1*)} \) and/or \( U_n^{(2*)} \) will be much more straightforward to implement practically.

The variability of the score in \( R_{11} \) and the effect of censoring can be seen in Figure 3.2. Note that the expected value of the score in \( R_{11} \) is the same for both cases.

In the \( d \)-dimensional case, the likelihood contribution that must be avoided in order to obtain a score test with regular normalisation is again the joint density

\[
\frac{\partial^d F(x_1, \ldots, x_d)}{\partial x_1 \ldots \partial x_d}
\]

but now in a region that extends to \((\infty, \ldots, \infty)\). Censoring the region where all margins exceed \( u \) achieves this. We remark for clarity that the modified tests developed here, like the Tawn (1988) and Ledford and Tawn (1996) tests, are tests of exact independence against positive association between extremes and, as such, typically reject independence when evaluated on asymptotically independent data. For further discussion of asymptotic independence and clarification of the behaviour of these tests in such cases see Proposition 2 of Ledford and Tawn (1996).
Chapter 3. Score tests of independence in multivariate extreme values

Figure 3.1: Histograms, with standard normal densities superimposed, and standard normal QQ-plots showing the empirical distributions of the score statistics $U_n^{(1*)}$, $U_n^{(2*)}$, $U_n^{(1)}$ and $U_n^{(2)}$ under independence of the marginal distributions. The sample size is $n = 300$ in each case, and the threshold that defines the boundaries of $R_{ij}$ is the 90% quantile of the unit Fréchet distribution, i.e. $u = -1/\log(0.9)$. The results shown were obtained through 20,000 repeated simulations.
Chapter 3. Score tests of independence in multivariate extreme values

Figure 3.2: Plots of the observed score contributions from region $R_{11}$ a) when $R_{11}$ is censored and b) when $R_{11}$ is not censored for sample size 2500 and thresholds $u = -1/\log(0.9)$ . In the first case the plotted points have mean 300 and empirical variance 3500; in the second the mean is 300 and the empirical variance is 10500. The discretisation in a) is because the score contributions here depend only on the number of points falling in $R_{11}$. The number of simulations was 400.

3.2.2 Evaluating the variances $\sigma_1^2$ and $\sigma_2^2$

In order to construct the test statistics given in Proposition 3.2, the variances $\sigma_1^2$ and $\sigma_2^2$, which depend implicitly on the value of the threshold $u$, are required. These variances have very complicated expressions and detailed numerical integration is needed in order to calculate their values accurately. Clearly, the harder it is to evaluate these variances for a given threshold $u$ then the more problematic it will be to implement the corresponding test statistics. To overcome this potential problem, we calculated $\sigma_1^2$ and $\sigma_2^2$ using accurate numerical quadrature for a range of $u$ values, and then examined several easily computed nonlinear approximations of the observed relationships.

For simplicity, we will work with the threshold probability $p = F(u) = \exp(-1/u)$ rather than directly with the threshold $u$ and will consider values of $p$ satisfying $p \in [0.75, 1 - \epsilon]$ for some small $\epsilon \geq 0$. Taking $\epsilon = 10^{-7}$, we found by polynomial regression that the standard deviation $\sigma_1$ is well approximated by the function

$$\sigma_1(p) = 1.463193 + 3.124356 \times 10^{-1} \log(-\log(1-p)) +$$

$$1.323150 \times 10^{-1} [\log(-\log(1-p))]^2 + 3.571265 \times 10^{-2} [\log(-\log(1-p))]^3;$$
and taking $\epsilon = 0$, that the relationship between $\sigma_2$ and $p$ is well approximated by
\[
\hat{\sigma}_2(p) = 1.107767 + 3.627841 \times 10^{-1}p - 8.438097 \times 10^{-2}p^2.
\]
The accurately calculated standard deviations together with these approximations are shown in Figure 3.3. The lines corresponding to the approximations and accurately calculated values are nearly coincident, indicating that the approximations provide high accuracy. Indeed, the maximum discrepancies that arise between the approximations and the accurately calculated standard deviations are $7 \times 10^{-4}$ for $\sigma_1$ and $1 \times 10^{-4}$ for $\sigma_2$, for $p$ within the respective range. Further analysis of the expressions for the variances shows that $\sigma_1 \to \infty$ and $\sigma_2 \to 2 \log 2$ as $u \to \infty$.

3.3 Comparison between regular and non-regular score tests

One of the most important features of the $R_{11}$ censoring procedure is that it yields tests with fast convergence. This results in the distribution of the normalised score statistic comparing well under independence, even for moderate sample sizes, with the standard normal distribution. In contrast, Figure 3.1 suggests that the unmodified score tests exhibit poor rates of convergence and a long upper tail which is due to the variability of the score in $R_{11}$, as can be seen in Figure 3.2. In this section we examine and compare several features of the original and modified tests in order to better quantify the benefits provided by our censoring approach.

Our first assessment is based on comparing critical values of the null distributions of the test statistics under the original and modified frameworks for sample sizes typical of applications. We undertake this via large scale simulation. Table 3.1 shows a range of critical values for the normalised scores based on $U_n^{(1)}$, $U_n^{(2)}$, $U_n^{(1*)}$ and $U_n^{(2*)}$ for threshold probabilities $p = 0.9$ and $p = 0.95$ and sample sizes $n = 1000$ and $n = 2500$. Comparing critical values under the modified and unmodified approaches, we see that the modified approach yields critical values that are always closer to the asymptotic values, and furthermore, that this benefit becomes more pronounced as increasingly extreme critical points are considered. There is also evidence that tests based on $U_n^{(1*)}$ have faster convergence than tests based on $U_n^{(2*)}$. We remark that more precise results
Chapter 3. Score tests of independence in multivariate extreme values

Figure 3.3: Graphs showing the variation of $\sigma_1$ and $\sigma_2$ as the threshold probability $p$ varies from $0.75$ to $1 - 10^{-7}$ and from $0.75$ to $1$, respectively, with the corresponding approximations superimposed (dotted line). Note that the two lines are virtually coincident.
may be obtained by choosing lower thresholds such as \( p = 0.85 \) or \( p = 0.8 \), and for clarity, that results for the Tawn model do not depend on the threshold.

Instead of considering the null distributions of the modified and unmodified tests, we now focus on their power functions. Intuition suggests that the likelihood censoring of our modified approach yields tests with reduced power. Although this is clearly detrimental, in practice we view this drawback as being offset by benefits such as increased applicability, regularity and overall ease-of-use. To assess the power function we again use simulation, but now obtain dependent Fréchet pairs using the scheme for generating from a bivariate extreme value distribution with logistic dependence outlined in Shi et al. (1992). In Figure 3.4 we compare the power functions of the different score tests for threshold probabilities \( p = 0.9 \) and \( p = 0.95 \) and sample sizes \( n = 1000 \) and \( n = 2500 \) using the 95% point of the standard normal distribution as the critical value of the test. Figure 3.5, which is derived from the results shown in Figure 3.4, depicts the actual loss of power associated with censoring region \( R_{11} \). Near independence, i.e. \( \alpha \approx 1 \), the test based on \( U^{(1)}_n \) has the highest power, and the test based on \( U^{(2)}_n \) is the next most powerful, as might be expected. The results suggest that the loss of power resulting from censoring region \( R_{11} \) is greatest for some \( \alpha < 1 \) and is of comparable magnitude to the loss of power between the unmodified Tawn (1988) and Ledford and Tawn (1996) tests. It is also clear that the tests examined here do not all have actual sizes that agree with their common nominal size (= 0.05), as they should do if the standard normal distribution was representative of their empirical law, e.g. the tests based on \( U^{(1)}_n \) and \( U^{(2)}_n \) have sizes bigger than 0.05. In order to make a comparison that overcomes this discrepancy, we ensure that the tests have the same actual sizes by using empirical critical values, as given in Table 3.1. The resulting power functions are depicted in Figure 3.6, and show that the loss of power resulting from censoring region \( R_{11} \) is now much smaller. The relationships between the nominal and actual sizes of the various tests are depicted in Figure 3.7. In contrast to the unmodified tests, the nominal and actual sizes for the modified tests show very close agreement.
Chapter 3. Score tests of independence in multivariate extreme values

<table>
<thead>
<tr>
<th>a)</th>
<th>1000</th>
<th>2500</th>
<th>10%</th>
<th>5%</th>
<th>2.5%</th>
<th>1%</th>
<th>0.5%</th>
<th>0.1%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10%</td>
<td>1.52</td>
<td>1.49</td>
<td>2.17</td>
<td>2.21</td>
<td>4.23</td>
<td>5.72</td>
<td>12.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.07)</td>
<td>(0.15)</td>
<td>(0.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>2.17</td>
<td>2.12</td>
<td>2.40</td>
<td>2.83</td>
<td>4.07</td>
<td>4.45</td>
<td>11.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.07)</td>
<td>(0.15)</td>
<td>(0.68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5%</td>
<td>4.23</td>
<td>2.83</td>
<td>4.07</td>
<td>5.45</td>
<td>5.81</td>
<td>6.35</td>
<td>12.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>5.72</td>
<td>4.07</td>
<td>5.45</td>
<td>6.35</td>
<td>7.01</td>
<td>7.57</td>
<td>14.73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.78)</td>
<td>(0.78)</td>
<td>(0.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5%</td>
<td>12.26</td>
<td>5.81</td>
<td>6.35</td>
<td>7.01</td>
<td>7.57</td>
<td>8.31</td>
<td>15.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.81)</td>
<td>(0.81)</td>
<td>(0.81)</td>
<td>(0.81)</td>
<td>(0.81)</td>
<td>(0.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2500.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5000.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10000.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Simulated and asymptotic critical values of normalised score statistics based on a) $U_n^{(1)}$, b) $U_n^{(2)}$, c) $U_n^{(1+)}$, and d) $U_n^{(2+)}$. Standard errors are given in parentheses. The simulation involved 5,000,000 replications of the normalised scores.
Chapter 3. Score tests of independence in multivariate extreme values

Figure 3.4: Power functions of normalised score statistics based on $U_n^{(1*)}$ (---), $U_n^{(1)}$ (· · ·), $U_n^{(2*)}$ (-----) and $U_n^{(2)}$ (---), based on the 95% asymptotic critical value. The threshold and sample size used were: a) $n = 1000$ and $p = 0.9$, b) $n = 1000$ and $p = 0.95$, c) $n = 2500$ and $p = 0.9$ and d) $n = 2500$ and $p = 0.95$. The number of simulations was 4,000,000.
Figure 3.5: The loss of power that results from censoring region $R_{11}$. The difference between the power of the Ledford and Tawn model and its modified version is represented by (——) and the difference for the Tawn model and its modified version is represented by (···). Cases a), b), c) and d) are as defined in Figure 3.4.
Figure 3.6: Power functions of normalised score statistics based on $U^{(1)}_n$ (---), $U^{(2)}_n$ (- - -), $U^{(3)}_n$ (---) and $U^{(4)}_n$ (- - -), based on the 95% empirical critical values. The threshold and sample size used were: a) $n = 1000$ and $p = 0.9$, b) $n = 1000$ and $p = 0.95$, c) $n = 2500$ and $p = 0.9$ and d) $n = 2500$ and $p = 0.95$. The number of simulations was 4,000,000.
Figure 3.7: Plot showing the actual and nominal sizes for normalised score statistics based on $U_n^{(1)}$ (---), $U_n^{(2)}$ (---) and $U_n^{(3)}$ (---) for a) $n = 1000$ and $p = 0.9$, b) $n = 1000$ and $p = 0.95$, c) $n = 2500$ and $p = 0.9$ and d) $n = 2500$ and $p = 0.95$. The grey line depicts the diagonal $y = x$. 
3.4 Further asymptotic properties of likelihood inference

In this section we examine the asymptotic distribution of the maximum likelihood estimator of the dependence parameter \( \alpha \) and consider the related issue of likelihood ratio tests under our modified framework. For a treatment of the behaviour of corresponding quantities under the unmodified framework, see Tawn (1988) and Ledford and Tawn (1996). We have the following:

**Theorem 3.3** For the likelihoods given in equations (3.6) and (3.7), under independence (i.e. \( \alpha = 1 \)) we have that \( \hat{\alpha} \) satisfies

\[
(1 - \hat{\alpha}) \sigma_1 \sqrt{n} \xrightarrow{d} Z \quad \text{and} \quad (1 - \hat{\alpha}) \sigma_2 \sqrt{n} \xrightarrow{d} Z
\]

as \( n \to \infty \), where the non-negative random variable \( Z \) has law

\[
\Pr(Z \leq x) = h(x) \Phi(x)
\]

for \( h(\cdot) \) the Heaviside step function and \( \Phi(\cdot) \) the standard normal distribution function.

**Proof.** By adopting similar arguments to Tawn (1988) or by applying Theorem 2 of Self and Liang (1987). ■

We consider now the asymptotic behaviour of the maximum likelihood estimator of the dependence parameter \( \alpha \) when the marginal distributions are no longer known to be unit Fréchet but are assumed asymptotically regular. Let \( \phi = (\phi^1, \ldots, \phi^g) \) be the vector of marginal parameters with true value \( \phi_0 \) and define the joint maximum likelihood estimator by \( (\hat{\alpha}, \hat{\phi}) \). Using a similar notation to that in Self and Liang (1987), let \( I(1, \phi_0) \) denote the expectation of \( n^{-1} I_n(1, \phi_0) \) with respect to the true joint density where \( -I_n(1, \phi_0) \) is the matrix of second derivatives with respect to \( \phi \) of the corresponding joint log-likelihood function evaluated at the point \( (\alpha, \phi) = (1, \phi_0) \), i.e. the matrix with entries \( \delta^2 \log L_n^{(k^j)}(1, \phi_0) / \partial \phi^i \partial \phi^j \) for \( i, j = 1, \ldots, g \) and \( k = 1 \) or \( 2 \).

**Theorem 3.4** Under independence and for multiparameter versions of the likelihoods given in equations (3.6) and (3.7), we have

\[
\left\{ (1 - \hat{\alpha}) \sigma_k \sqrt{n}, (\hat{\phi} - \phi_0)/\sqrt{n} \right\} \xrightarrow{d} (Z, Z_1, \ldots, Z_g) \quad \text{for } k = 1, 2
\]
as \( n \to \infty \), where \( Z \) is as defined in equation (3.8) and \( Z_1, \ldots, Z_q \) are zero mean normal random variables. The covariance matrix of the random vector is block-diagonal with non-null entries \( \{1, I^{-1}(\phi_0)\} \).

**Proof.** This result is obtained by applying Theorem 2 of Self and Liang (1987).

**Theorem 3.5** Let \( L_n^{(k*)}(\hat{\alpha}) \) for \( k = 1, 2 \) denote the joint maximum of the appropriate multiparameter likelihood taken over the dependence parameter \( \alpha \in (0,1] \) and the marginal parameters, and \( L_n^{(k*)}(1) \) the corresponding maximum taken over the marginal parameters under the constraint \( \alpha = 1 \). Then

\[
2 \log \left\{ \frac{L_n^{(k*)}(\hat{\alpha})}{L_n^{(k*)}(1)} \right\} \xrightarrow{\text{w}} Z^2 \quad \text{as } n \to \infty
\]

where \( Z \) is as defined in equation (3.8).

**Proof.** Follows from Self and Liang (1987).

To examine informally the rate of convergence of this latter result and its analogue for the unmodified tests (see Tawn, 1988, and Ledford and Tawn, 1996) we compare the empirical distribution of the likelihood ratio statistics for a set of 300 independent points and the distribution of the variable \( Z^2 \). Results are shown in Figure 3.8. Kolmogorov-Smirnov tests were also used and the test statistic values obtained when using the likelihoods \( L_n^{(1*)} \), \( L_n^{(2*)} \), \( L_n^{(1)} \) and \( L_n^{(2)} \) were 0.01, 0.02, 0.06 and 0.10, respectively. From the analysis of Figure 3.8 and since values of the Kolmogorov-Smirnov statistic closer to zero corresponds to distributions with closer agreement, we conclude that convergence to \( Z^2 \) is more rapid for the modified tests. The power functions of the associated tests were also examined through simulation (not reported here). Our findings suggest that the power functions associated with these tests behave similarly to those of the score tests.

### 3.5 Alternative dependence models

Throughout this chapter we have used the logistic dependence model to derive test statistics and also to examine the behaviour of these test statistics under departures from the null model. It is important to assess the sensitivity of these results to this
Figure 3.8: Histograms, with the density of $Z^2$ superimposed, and QQ-plots of simulated values of the likelihood ratio statistic using $L_n^{(2)}$, $L_n^{(3)}$, $L_n^{(4)}$ and $L_n^{(5)}$ for samples with size $n = 300$. The threshold probability used was $p = 0.9$ and the number of simulations was 20,000.
particular choice of dependence model, both in terms of the structure of the tests and their behaviour. To achieve this we repeated the derivation of modified score statistics using the mixed model, and additionally, examined the behaviour of the test statistics obtained previously for the logistic model using dependent data simulated from the mixed model, and vice-versa.

In this case, using the dependence function \( V(x, y) = x^{-1} + y^{-1} - \theta(x + y)^{-1} \) for \( 0 \leq \theta \leq 1 \), the score contributions for each region \( R_{kl} \) are of the form

\[
S_{0}^{(1m)}(x, y) = (x + y)^{-1} - (x^2 + y^2)(x + y)^{-2} + 2(xy)^2(x + y)^{-3} \quad \text{if} \quad (x, y) \notin R_{11},
\]

\[
S_{1}^{(1m)}(x, y) = (2u)^{-1} \exp(-2u^{-1})/ \{1 - 2 \exp(-u^{-1}) + \exp(-2u^{-1})\} \quad \text{if} \quad (x, y) \in R_{11},
\]

using the scheme of censoring only in region \( R_{11} \), and

\[
S_{00}^{(2m)}(x, y) = (2u)^{-1},
\]

\[
S_{01}^{(2m)}(x, y) = (u + y)^{-1} - y^2(u + y)^{-2},
\]

\[
S_{10}^{(2m)}(x, y) = (x + u)^{-1} - x^2(x + u)^{-2},
\]

\[
S_{11}^{(2m)}(x, y) = (2u)^{-1} \exp(-2u^{-1})/ \{1 - 2 \exp(-u^{-1}) + \exp(-2u^{-1})\},
\]

using the multiple censoring scheme.

The mixed model yields likelihood ratio tests that are identical to those for the logistic model, and the resulting score tests and behaviour of the maximum likelihood estimator of the dependence parameter are identical to those for the logistic model up to a scaling constant, except that \((1 - \hat{\alpha})\) is replaced by \(\hat{\theta}\) because independence in the mixed model occurs when \(\theta = 0\). Simulation based results that examine the behaviour of the logistic test statistic evaluated on dependent data from the mixed model, and vice-versa, suggest that the test statistics are properly able to detect departures from independence that are outside the underlying parametric family from which they were derived. The similarity of structure and behaviour that we observe here suggests that our results are reasonably generic features of independence testing within extremes and are fairly robust.

### 3.6 Conclusions

The proposed regularised score tests have been shown to have good performance in practice and to overcome the convergence difficulties encountered when using the existing non-regular score tests. Our test statistics converge rapidly to standard normal
Chapter 3. Score tests of independence in multivariate extreme values

distributions and are consequently more useful in practical situations with realistic amounts of data. However, this benefit is at the expense of some reduction in power, but this loss of power has been shown to be of an acceptable size. In contrast to the Tawn (1988) and the Ledford and Tawn (1996) score tests, our new score tests have close agreement between their actual and nominal sizes. Similar results and conclusions hold for tests based on the likelihood ratio statistic.

There is another model that we could compare with the previous ones: the one resulting from censoring the data in the whole plane according to the four regions used previously. Obviously, the only information such a censoring scheme conveys is the number of points which fall into each region. The likelihood contribution given by an observation in \( R_{kl} \) is then just the probability of falling in that region, i.e. \( F(u, u) \) for \( R_{11} \), \( F(u) - F(u, u) \) for \( R_{01} \) and \( R_{10} \), and \( F(u, u) = \Pr(X > u, Y > u) \) for \( R_{11} \). Naturally, this results in a regular score test too, and it can be shown that the variance of the resulting score for the logistic dependence structure converges to \( 4 \log^2 2 \) as \( n \to \infty \). This methodology can be used to underpin a contingency table type test of independence.
Chapter 4

A pseudo-polar representation of asymptotic independence

A fundamental and general result in classical multivariate extreme value (MEV) theory is the pseudo-polar representation of the MEV distribution in terms of a positive angular measure that is essentially arbitrary apart from having to satisfy some simple normalisation conditions (see equation (1.11) in Chapter 1). Concentrating in this chapter on the bivariate case, for simplicity, an analogue of this result will be shown to hold for a much wider class of tail models. Our results are similar to the MEV ones but yield a different set of normalisation conditions that may be seen to extend the existing pseudo-polar representation for bivariate extremes to encompass asymptotically independent tails. A constructional procedure for obtaining parametric asymptotically independent joint tail models is then developed and some examples are also given.

4.1 Introduction

It was mentioned in Chapter 1, Section 1.5.3, that both standard parametric and non-parametric estimation techniques present some problems in the particular case where data are asymptotically independent, i.e. when their distribution is in the domain of attraction of independence. Indeed, it is sometimes the case that MEV models are not very useful in such cases because they are based on componentwise maxima and consequently may focus on observations that tend never to arise in any sample.
However, asymptotic independence is an important case in practice both for applications and for theoretical development, see e.g. de Haan and Ronde (1998) or Ledford and Tawn (1997), and arises for most classical families of distribution, as listed in Capéraà et al. (2000) and Heffernan (2000).

Ledford and Tawn (1996, 1997) proposed a bivariate joint tail model that included the asymptotically independent case. In their approach, the survivor function of a bivariate random variable \((X, Y)\) with unit Fréchet margins is assumed to satisfy

\[
\Pr(X > x, Y > y) = F_{XY}(x, y) = \frac{L(x, y)}{x^{c_1}y^{c_2}}
\]

where \(L\) is a bivariate slowly varying (BSV) function and \(c_1 + c_2 = \eta^{-1}\) where \(\eta \in (0, 1)\) is the coefficient of tail dependence. This structure results in a characterisation of asymptotic independence and includes asymptotic dependence and exact independence as special cases. It also provides a smooth transition between perfect dependence and exact independence, and allows negative dependence between the marginal extremes, yielding a very flexible and broadly applicable framework for statistical modelling.

Let \(g\) denote the limit function of \(L\), so that for all \((x, y) \in \mathbb{R}^2_+\) and \(c > 0\)

\[
g(x, y) = \lim_{r \to \infty} \frac{L(rx, ry)}{L(r, r)} \quad \text{and} \quad g(cx, cy) = g(x, y).
\]

Using this result, the limit function \(g\) can be shown to be constant along any ray \(y = ax\) for \(a > 0\), and so it can be written as \(g(x, y) = g_*(\{x/(x + y)\}) = g_*(w)\) for \(w = x/(x + y) \in (0, 1)\). This \(g\) measures the asymptotic ray dependence of the BSV function \(L\), and \(L\) can be defined as asymptotically ray dependent if \(g_*(w)\) varies with \(w\) and as asymptotically ray independent if \(g_*(w)\) is constant over different rays.

For this reason, \(g\) is also called the ray dependence function.

A bivariate-threshold-based model based on the above was developed and applied by Ledford and Tawn (1997). Their approach focused on modelling \(L\) as a mixture of ray independence and ray dependence. A submodel of this was used by Bruun and Tawn (1998) in modelling offshore storm data. However, although this model has been used in practice, in some special cases it can be shown that the joint density obtained is valid only when \(w = x/(x + y)\) is in some restricted range of \((0, 1)\), and negative values of the density may arise if \(w\) becomes too close to 0 or 1, see Ledford (1996), Section 7.2. The new model suggested here overcomes this problem and ensures that we obtain valid densities always.
4.2 Modelling dependence within bivariate joint tails

Consider a bivariate random variable \((X, Y)\) with unit Fréchet marginal distributions that satisfies equation (4.1). The aim here is to estimate the joint tail of the distribution function \(F_{XY}\), which is the same as modelling the extremal dependence structure, since the margins are assumed known.

Let \(u\) denote a high threshold. We consider the behaviour of the bivariate conditional random variable \((S, T)\) defined by

\[
(S, T) = \lim_{u \to \infty} (X/u, Y/u) \mid (X > u, Y > u),
\]

in the sense that, for all \((s, t) \in [1, \infty) \times [1, \infty),\)

\[
Pr(S > s, T > t) = \lim_{u \to \infty} \frac{Pr(X > su, Y > tu)}{Pr(X > u, Y > u)}.
\]  (4.4)

Examining now the distributional (i.e. weak) convergence in equation (4.4), we have that

\[
Pr(S > s, T > t) = \lim_{u \to \infty} \frac{\mathcal{L}(us, ut)}{\mathcal{L}(u, u)s^\alpha t^\beta} = \frac{g(s,t)}{s^\alpha t^\beta},
\]

for \(g\) as in equation (4.2). This survivor function can be simplified when \(c_1 = c_2 = (2\eta)^{-1}\) to the form

\[
Pr(S > s, T > t) = \frac{g(s,t)}{(st)^{1/(2\eta)}},
\]  (4.5)

Note that the condition \(c_1 = c_2 = (2\eta)^{-1}\) results in no loss of generality here since asymmetry can always be absorbed by the limit function \(g\).

This limit function \(g\), or more specifically its pseudo-polar representation \(g_*(w)\), where \(w = s/(s + t)\), can be geometrically interpreted as follows: since the survivor function of \((S, T)\) can be written in the form

\[
\overline{F}_{ST}(s, t) = \frac{g_*(w)}{(s + t)^{1/\eta} \{w(1 - w)\}^{1/(2\eta)}},
\]

then the quantity \(g_*(w) / \{w(1 - w)\}^{1/(2\eta)}\) is proportional to the profile of the survivor function along contours \(s + t = k\) where \(k\) is a constant.

Our objective now is to obtain a parametric expression for the survivor function of \((S, T)\) in equation (4.5) that will provide a model for the joint tail of \(F_{XY}\). However, the task of choosing an appropriate \(g\) (or \(g_*\)) is not immediate since \(g\) has to be chosen so that
Chapter 4. A pseudo-polar representation of asymptotic independence

A proper joint density is obtained for the variable \((S, T)\). We address this issue via a different approach which is described below.

Transforming to the pseudo-radial and angular coordinates defined by \(R = S + T\) and \(W = S/R\), and assuming that the limiting function \(g_*\) is twice differentiable then the density of \((R, W)\) exists and can be shown to satisfy

\[
f(r, w) = r^{-(1+1/\eta)}h_\eta(w),
\]

for \(w \in (0,1)\) and \(r \in [\max\{1/w, 1/(1-w)\}, \infty)\), where the function \(h_\eta\) is a non-negative measure density\(^1\) on \((0,1)\) determined by \(g_*\) and \(\eta\)\(^2\). Basically our approach is to look for a valid function \(h_\eta\) for equation (4.6), and then use this density to work back (by integrating) to obtain a joint survivor function for \((S, T)\). A key part of our analysis is establishing the regularity conditions which ensure \(h_\eta\) is a valid function.

Thus, as suggested above, we consider reconstructing the survivor function of \((S, T)\), \(F_{ST}\), from the density in equation (4.6). Letting \(r^* = \max\{s/t, t/(1-w)\}\), we have

\[
\Pr(S > s, T > t) = \int_{w=0}^{1} \int_{r=r^*}^{\infty} r^{-(1+1/\eta)}h_\eta(w) \, dr \, dw
\]

\[
= \eta \int_{0}^{1} \left\{ \min\left(\frac{w}{s}, \frac{1-w}{t}\right) \right\} \frac{1}{\eta} h_\eta(w) \, dw
\]

\[
= \eta \int_{0}^{\frac{1}{s+t}} \left(\frac{w}{s}\right)^{1/\eta} h_\eta(w) \, dw + \eta \int_{\frac{1}{s+t}}^{1} \left(\frac{1-w}{t}\right)^{1/\eta} h_\eta(w) \, dw,
\]

for \((s,t) \in [1, \infty) \times [1, \infty)\). Providing \(h_\eta\) is known, this representation can be used to obtain parametric models for the joint survivor function of \((S, T)\) and hence models for \(g_*\), and consequently models for the joint tail of the survivor function \(F_{XY}\).

In order to derive the required regularity conditions on \(h_\eta\), we write \(s = t = t_0\) in equation (4.7) and exploit equation (4.2)\(^3\) to obtain

\[
\eta^{-1} = \int_{0}^{1/2} w^{1/\eta} h_\eta(w) \, dw + \int_{1/2}^{1} (1-w)^{1/\eta} h_\eta(w) \, dw.
\]

We refer to this equation as the normalisation condition. Note the similarity to the condition in equation (1.11) for the standard bivariate extreme value (BEV) case.

\(^1\)Function \(h_\eta\) is such that \(h_\eta(w) = \frac{dH_\eta(w)}{dw}\) if the measure \(H_\eta\) is differentiable; atomic masses can be considered otherwise.

\(^2\)\(h_\eta(w) = \{w(1-w)^{-1/2(\eta-1)} [g_*(w)/(4\eta^2)] + g_*(w) \{w(1-w)\} (2w - 1) - g_*(w) \{w(1-w)\}^2\}.

\(^3\)From equation (4.2), \(g(1,1) = g_*(1/2) = 1\).
We have the following result:

**Theorem 4.1** Let \((X, Y)\) satisfy equation (4.1) so that

\[
\lim_{u \to \infty} \Pr(X > ux, Y > uy \mid X > u, Y > u) = g_*(x/(x+y))/(xy)^{1/(2n)}
\]

(4.8)

for \((x, y) \in [1, \infty) \times [1, \infty)\), where \(g_*\) is the limit function defined following equation (4.2).

Then, for \(w = x/(x+y)\), \(g_*(w)\) satisfies

\[
g_*(w) = \eta \left(\frac{1-w}{w}\right)^{1/(2n)} \int_0^w z^{1/\eta} h_\eta(z) \, dz + \eta \left(\frac{w}{1-w}\right)^{1/(2n)} \int_w^1 (1-z)^{1/\eta} h_\eta(z) \, dz,
\]

(4.9)

where \(h_\eta\) is a non-negative measure density on \([0, 1]\) satisfying

\[
\eta^{-1} = \int_0^{1/2} w^{1/\eta} h_\eta(w) \, dw + \int_{1/2}^1 (1-w)^{1/\eta} h_\eta(w) \, dw.
\]

(4.10)

Conversely, given any \(h_\eta\) satisfying equation (4.10), then equations (4.9) and (4.8) define a valid joint survivor function for the limiting random variable in equation (4.8).

**Proof.** (\(\Rightarrow\)) Immediate since equation (4.9) follows from equations (4.7) and (4.8).

(\(\Leftarrow\)) We need to check that for \(h_\eta\) satisfying (4.10), then equations (4.8) and (4.9) together define a joint valid survivor function, i.e. the associated joint density is non-negative everywhere and also integrates to one.

Non-negativity is obvious from equation (4.6), since the density of \((R, W)\) is given by \(r^{-1+1/\eta} h_\eta(w) > 0\) for all \(r\) and \(w \in (0, 1)\). To verify that the density integrates to one over the appropriate domain, note that, by condition (4.10),

\[
\int_0^1 \int_\max\{w,1-w\}^\infty f(r, w) \, dr \, dw = \eta \int_0^{1/2} w^{1/\eta} h_\eta(w) \, dw + \eta \int_{1/2}^1 (1-w)^{1/\eta} h_\eta(w) \, dw = 1.
\]

The above result shows that, given \(\eta \in (0, 1]\), any non-negative measure density \(h_\eta\) on \([0, 1]\) satisfying the normalisation condition (4.10) provides a joint survivor function for \((S, T)\) via the construction detailed above, and thus yields a valid model for the joint tail of \(F_{XY}\).
To exploit this within applications we follow the usual approach of treating a limit as the basis for an approximation and so obtain the following joint tail model for the original bivariate variable $(X, Y)$

$$F_{XY}(x, y) = \lambda F_{ST}(x/u, y/u)$$ (4.11)

for $x > u$ and $y > u$, where $\lambda = \Pr(X > u, Y > u)$ and $u$ is a high threshold.

### 4.3 Examples of parametric models

Some illustrative examples of joint tail parametric models now follow. Most of them were obtained from the parametric families of BEV distributions in Section 1.4.4 by modifying their dependence structure. Let $w = s/(s + t)$ throughout.

**Example A:** A joint tail model based on a modification of the logistic dependence structure.

Consider the measure density defined by

$$h_\eta(w) = A_{\alpha \eta} \left\{ w^{-1/\alpha} + (1 - w)^{-1/\alpha} \right\}^{\alpha/\eta - 2} \left\{ w(1 - w) \right\}^{-1 - 1/\alpha},$$

where

$$A_{\alpha \eta} = \frac{\eta - \alpha}{\alpha \eta^2 (2 - 2\alpha/\eta)}$$

for the coefficient of tail dependence $\eta \in [0, 1]$ and the dependence parameter $\alpha > 0$.

Elementary integration shows that $h_\eta$ as specified above satisfies the normalisation condition (4.10). Working through the construction of equation (4.7) yields

$$F_{ST}(s, t) = \frac{1}{2 - 2\alpha/\eta} \left\{ s^{-1/\eta} + t^{-1/\eta} - \left( s^{-1/\alpha} + t^{-1/\alpha} \right)^{\alpha/\eta} \right\}$$ (4.12)

for $(s, t) \in [1, \infty) \times [1, \infty)$. The associated limit function $g$ is given by

$$g(s, t) = (st)^{1/(2\eta)} F_{ST}(s, t),$$

which, on writing $w = s/(s + t)$, may be written as

$$g(s, t) = g_s(w) = \frac{(w(1 - w))^{1/(2\eta)}}{2 - 2\alpha/\eta} \left[ w^{-1/\eta} + (1 - w)^{-1/\eta} - \left( w^{-1/\alpha} + (1 - w)^{-1/\alpha} \right)^{\alpha/\eta} \right].$$ (4.13)
Chapter 4. A pseudo-polar representation of asymptotic independence

Plots of \( \log g_\alpha(w) \) are shown in Figure 4.1 for different values of \( \alpha \) and \( \eta \). It can be seen from this figure that the graph is concave when \( \alpha < 2\eta \), which corresponds to \( \mathcal{L} \) having a limit function which exhibits concave ray dependence. The graph is flat when \( \alpha = 2\eta \), which corresponds to ray independence, and is convex when \( \alpha > 2\eta \), which corresponds to convex ray dependence. It can also be verified that \( g_\alpha(w) \) converges to 1 for all \( w \in (0, 1) \) as \( \alpha \) tends to infinity. That is, ray independence arises for \( g_\alpha \) so that \( g_\alpha(w) \equiv 1 \) for all \( w \in (0, 1) \), when \( \alpha = 2\eta \) or in the case \( \alpha \to \infty \).

A geometrical interpretation of \( g_\alpha \) is obtained by studying the quantity

\[
\frac{g_\alpha(w)}{w(1-w)^{1/(2\eta)}} = \left( 2 - 2^{\alpha/\eta} \right)^{-1} \left[ w^{-1/\eta} + (1-w)^{-1/\eta} - \left( w^{-1/\alpha} + (1-w)^{-1/\alpha} \right)^{\alpha/\eta} \right].
\]

Plots of the logarithm of this quantity are shown in Figure 4.2. The resulting shapes unlike those for \( g_\alpha \), are not always either concave or convex.

Now, using the approximation given by equation (4.11), the parametric model for the joint survivor function of \((S, T)\) yields the following model for the original bivariate \((X, Y)\) variable:

\[
\overline{F}_{XY}(x, y) = \frac{\lambda u^{1/\eta}}{2 - 2^{\alpha/\eta}} \left\{ x^{-1/\eta} + y^{-1/\eta} - \left( x^{-1/\alpha} + y^{-1/\alpha} \right)^{\alpha/\eta} \right\}
\]

for \((x, y) \in [u, \infty) \times [u, \infty)\), where \(u\) is a chosen high threshold and \(\lambda\) is the joint threshold exceedance probability.

Clearly, the survivor functions \(\overline{F}_{ST}\) and \(\overline{F}_{XY}\) are bivariate regularly varying functions of index \(-1/\eta\).

**Note:** The modified logistic measure density given above does not necessarily have finite mass. For example, the integral of \(h_\eta\) over \(w \in (0, 1)\) does not converge when \(\alpha \geq \eta\), although it does converge when \(\alpha < \eta\). This demonstrates that the normalisation condition (4.10) admits a characteristically different class of angular measures than that admitted by the standard BEV theory.

Note also that unlike in the standard BEV case the dependence parameter \(\alpha\) may be greater than 1 since the density function

\[
f_{XY}(x, y) = \frac{\lambda u^{1/\eta} (\eta - \alpha)}{2\eta^2 (2 - 2^{\alpha/\eta})} \left( x^{-1/\alpha} + y^{-1/\alpha} \right)^{\alpha/\eta - 2} (xy)^{-1+1/\alpha}, \quad x, y > u
\]

for the variable \((X, Y)\) remains valid even when \(\alpha > 1\), for every value of \(\eta \in (0, 1]\). This dependence parameter \(\alpha\) measures the dependence between the variables for a fixed \(\eta\).
Figure 4.1: Plots of $\log g_w(w)$ as in Example A: top) for fixed $a = 0.5$ (note that as a function of $w$ the graph is convex when $\eta < 0.25$, is concave when $\eta > 0.25$ and is constant when $\eta = 0.25$), and bottom) for fixed $\eta = 0.5$ (note that the graph is concave when $\alpha < 1$, convex when $\alpha > 1$ and constant when $\alpha = 1$). Interestingly, the bottom figure flattens as a function of $w$ as $\alpha$ increases.
Figure 4.2: Plots of \( \log \left[ g_\alpha(w)/\{w(1-w)\}^{1/(2\eta)} \right] \) as in Example A: top) for fixed \( \alpha = 0.5 \) and bottom) for fixed \( \eta = 0.5 \). The figures on the right are enlargements of the figures on the left for particular values of \( \eta \) or \( \alpha \), showing that the graph is neither convex nor concave.
We remark also that although \( \alpha = \eta \) gives degenerate behaviour in \( A_{\alpha \eta} \), the limit of \( A_{\alpha \eta} \) as \( \alpha \) converges to \( \eta \) is finite, and it is straightforward to show that \( \lim_{\alpha \to \eta} A_{\alpha \eta} = (2\eta^2 \ln 2)^{-1} \). The survivor model for the variable \((S, T)\) in this special case can be treated as the limit of \( \bar{F}_{ST} \) as \( \alpha \to \eta \) and yields the following non-degenerate survivor function

\[
\lim_{\alpha \to \eta} \bar{F}_{ST}(s, t) = \frac{(s^{-1/\eta} + t^{-1/\eta}) \log (s^{-1/\eta} + t^{-1/\eta}) - s^{-1/\eta} \log (s^{-1/\eta}) - t^{-1/\eta} \log (t^{-1/\eta})}{2 \log 2}.
\]

The survivor model for the original variable \((X, Y)\) is obtained in the same way.

Due to the simplicity and flexibility of this logistic-type tail model it will be one of the main parametric models used in our later applied work.

An alternative parameterisation of Example A may be sometimes of use, and may be obtained from the measure density

\[
h_{\eta}(w) = \frac{1 - \alpha'}{\alpha' \eta^2(2 - 2\alpha')} \left\{ w^{-1/(\eta \alpha')} + (1 - w)^{-1/(\eta \alpha')} \right\}^{\alpha' - 2} \left\{ w(1 - w)^{-1/(\eta \alpha')} \right\}
\]
for \( \eta \in (0, 1] \) and dependence parameter \( \alpha' > 0 \), where \( \alpha' = \alpha/\eta \) for \( \alpha \) and \( \eta \) as given previously.

**Example B:** A joint tail model based on a modification of the asymmetric logistic dependence structure.

First, we present a modified version of the asymmetric logistic model using a parameterisation based on that of the standard asymmetric logistic model. Since equation (4.10) says nothing about the end behaviour of a density \( h_{\eta} \) at positions \( w = 0 \) and \( w = 1 \), masses at these end points may be chosen entirely arbitrary. However, although they do not affect the survivor functions \( \bar{F}_{ST} \) and \( \bar{F}_{XY} \), end masses are presented here as they will prove useful later, in Chapter 6. The particular form of these end masses were chosen so that the results obtained in Chapter 6 have a more direct relationship with the existing standard BEV analogues. Finally, an alternative parameterisation is considered that offers benefits for estimation and applications.

Consider the measure with density

\[
h_{\eta}(w) = D_{\alpha \eta \phi} \left\{ \left( \frac{w}{\theta} \right)^{-1/\alpha} + \left( \frac{1 - w}{\phi} \right)^{-1/\alpha} \right\}^{\alpha/\eta - 2} \left\{ \frac{w}{\theta} \left( \frac{1 - w}{\phi} \right)^{-(1+1/\alpha)} \right\} \quad (4.15)
\]
for $0 < w < 1$ and with atoms of mass $(1 - \phi^{1/\eta})/(\eta N_{\theta \phi})$ and $(1 - \theta^{1/\eta})/(\eta N_{\theta \phi})$ at $w = 0$ and $w = 1$, respectively and where

$$D_{\alpha \eta \theta \phi} = \frac{\eta - \alpha}{\alpha \eta^2 \theta \phi N_{\theta \phi}}$$

and

$$N_{\theta \phi} = \theta^{1/\eta} + \phi^{1/\eta} - (\theta^{1/\alpha} + \phi^{1/\alpha})^{\alpha/\eta}$$

for the dependence parameters $\theta \in [0, 1]$, $\phi \in [0, 1]$, $\eta \in (0, 1]$ and $\alpha > 0$.

It is straightforward to show that the normalising condition (4.10) is verified for $h_\eta$ as specified in equation (4.15). Thus, by equation (4.7), this measure density yields the joint survivor model

$$F_{\text{ST}}(s, t) = N_{\theta \phi}^{-1} \left[ \left( \frac{s}{\theta} \right)^{-1/\eta} + \left( \frac{t}{\phi} \right)^{-1/\eta} - \left\{ \left( \frac{s}{\theta} \right)^{-1/\alpha} + \left( \frac{t}{\phi} \right)^{-1/\alpha} \right\}^{\alpha/\eta} \right] , \quad (4.16)$$

where $(s, t) \in [1, \infty) \times [1, \infty)$. The associated ray dependence function $g_*$ is then given by

$$g_*(w) = \frac{\{w(1-w)\}^{1/(2\eta)}}{N_{\theta \phi}} \left[ \left( \frac{w}{\theta} \right)^{-1/\eta} + \left( \frac{1 - w}{\phi} \right)^{-1/\eta} - \left\{ \left( \frac{w}{\theta} \right)^{-1/\alpha} + \left( \frac{1 - w}{\phi} \right)^{-1/\alpha} \right\}^{\alpha/\eta} \right]$$

and equation (4.11) provides the following joint survivor model:

$$F_{XY}(x, y) = \frac{\lambda u^{1/\eta}}{N_{\theta \phi}} \left[ \left( \frac{x}{\theta} \right)^{-1/\eta} + \left( \frac{y}{\phi} \right)^{-1/\eta} - \left\{ \left( \frac{x}{\theta} \right)^{-1/\alpha} + \left( \frac{y}{\phi} \right)^{-1/\alpha} \right\}^{\alpha/\eta} \right] \quad (4.17)$$

for $(x, y) \in [u, \infty) \times [u, \infty)$, where $u$ is a chosen high threshold and $\lambda$ is the joint threshold exceedance probability.

Similarly to Example A, the case $\alpha = \eta$ produces degenerate behaviour in $D_{\alpha \eta \theta \phi}$ and so is taken to be the limiting case as $\alpha \to \eta$ \footnote{\[ \lim_{\alpha \to \eta} D_{\alpha \eta \theta \phi} = \left[ (\phi^{1/\eta} + \theta^{1/\eta}) \log \left( \phi^{1/\eta} + \theta^{1/\eta} \right) - \theta^{1/\eta} \log \left( \theta^{1/\eta} \right) - \phi^{1/\eta} \log \left( \phi^{1/\eta} \right) \right]^{-1}. \]}. A proper survivor function for $(S, T)$ is obtained for this case taking the limit of (4.16) as $\alpha \to \eta$ \footnote{\[ \lim_{\alpha \to \eta} F_{\text{ST}}(s, t) = \frac{\left( (s^{1/\eta} + t^{1/\eta}) \log \left( s^{1/\eta} + t^{1/\eta} \right) - s^{1/\eta} \log \left( s^{1/\eta} \right) - t^{1/\eta} \log \left( t^{1/\eta} \right) \right)}{s^{1/\eta} \log \left( s^{1/\eta} \right) + t^{1/\eta} \log \left( t^{1/\eta} \right) - (s^{1/\eta} + t^{1/\eta})^{1/\eta} \log \left( s^{1/\eta} + t^{1/\eta} \right) \log \left( s^{1/\eta} + t^{1/\eta} \right)}. \]}. The limit of (4.17) as $\alpha \to \eta$ also results in a proper joint tail model for the original bivariate $(X, Y)$ variable.

An alternative parameterisation:

The above models can be simplified from four to three parameters by writing $\varrho = \phi/\theta$, so defining $N_\varrho = 1 + \varrho^{1/\eta} - (1 + \varrho^{1/\alpha})^{\alpha/\eta}$, the survivor models in equations (4.16) and (4.17) can be reduced to the following models:

$$F_{\text{ST}}(s, t) = N_\varrho^{-1} \left[ s^{-1/\eta} + \left( \frac{s}{\varrho} \right)^{-1/\eta} - \left\{ s^{-1/\alpha} + \left( \frac{s}{\varrho} \right)^{-1/\alpha} \right\}^{\alpha/\eta} \right] \quad (4.18)$$

\footnote{\[ \lim_{\alpha \to \eta} D_{\alpha \eta \theta \phi} = \left[ (\varrho^{1/\eta} + \theta^{1/\eta}) \log \left( \varrho^{1/\eta} + \theta^{1/\eta} \right) - \theta^{1/\eta} \log \left( \theta^{1/\eta} \right) - \varrho^{1/\eta} \log \left( \varrho^{1/\eta} \right) \right]^{-1}. \]}

\footnote{\[ \lim_{\alpha \to \eta} F_{\text{ST}}(s, t) = \frac{\left( (s^{1/\eta} + t^{1/\eta}) \log \left( s^{1/\eta} + t^{1/\eta} \right) - s^{1/\eta} \log \left( s^{1/\eta} \right) - t^{1/\eta} \log \left( t^{1/\eta} \right) \right)}{s^{1/\eta} \log \left( s^{1/\eta} \right) + t^{1/\eta} \log \left( t^{1/\eta} \right) - (s^{1/\eta} + t^{1/\eta})^{1/\eta} \log \left( s^{1/\eta} + t^{1/\eta} \right) \log \left( s^{1/\eta} + t^{1/\eta} \right)}. \]}.
and
\[
\bar{F}_{XY}(x, y) = \frac{\lambda u^{1/\eta}}{N_{u}} \left[ x^{-1/\eta} + \left( \frac{y}{\theta} \right)^{-1/\eta} - \left\{ x^{-1/\alpha} + \left( \frac{y}{\theta} \right)^{-1/\alpha} \right\}^{\alpha/\eta} \right] \tag{4.19}
\]
for \( \varrho > 0 \). Clearly, these are the versions we should use for modelling since the dependence parameters \( \theta \) and \( \phi \) in the previous models are not separately identifiable.

Similarly, under this parameterisation, the ray dependence function has the simplified form
\[
g_s(w) = \frac{\{w(1-w)\}^{1/2n}}{N_{u}} \left[ w^{-1/\eta} + \left( \frac{1-w}{\varrho} \right)^{-1/\eta} - \left\{ w^{-1/\alpha} + \left( \frac{1-w}{\varrho} \right)^{-1/\alpha} \right\}^{\alpha/\eta} \right].
\]

It is easy to check that, as in the previous example, \( g_s(w) \equiv 1 \) for all \( w \in (0,1) \) both when \( \alpha = 2\eta \), and when \( \alpha \to \infty \) independently of the value of the parameter \( \varrho \). It is also apparent that \( g_s \) is concave ray dependent when \( \alpha < 2\eta \) and is convex ray dependent when \( \alpha > 2\eta \), as can be seen from Figure 4.3. This figure also shows the behaviour of \( g_s \) as \( \varrho \) varies.

The profile of the survivor function \( \bar{F}_{ST} \) as described in equation (4.18) is given by
\[
\frac{g_s(w)}{\{w(1-w)\}^{1/2n}} = N_{u}^{-1} \left[ w^{-1/\eta} + \left( \frac{1-w}{\varrho} \right)^{-1/\eta} - \left\{ w^{-1/\alpha} + \left( \frac{1-w}{\varrho} \right)^{-1/\alpha} \right\}^{\alpha/\eta} \right].
\]

Plots of the logarithm of this quantity (not reported) are similar to those for Example A, except for the asymmetry factor.

The measure density in (4.15) also has a simplified version for \( w \in (0,1) \), that is
\[
h_{\eta}(w) = \frac{\eta - \alpha}{\alpha \eta^2 \varrho N_{u}} \left\{ w^{-1/\alpha} + \left( \frac{1-w}{\varrho} \right)^{-1/\alpha} \right\}^{\alpha/\eta - 2} \left( \frac{w}{\varrho} \right)^{(1+1/\alpha)},
\]
but the same method of simplification cannot be applied to the mass at \( w = 0 \) and \( w = 1 \). Therefore, and since these end masses do not have any effect in the survivor models \( \bar{F}_{ST} \) and \( \bar{F}_{XY} \), they can be discarded here. As above, the special case \( \alpha = \eta \) is taken to be the limit as \( \alpha \to \eta \).

Our simplified parameterisation works fine for modelling the joint tail of \( (S,T) \) and thus of \( (X,Y) \) and also for applied statistical modelling based on this case. However complications arise when the applications in Chapter 6 are considered and it is then easier to use the four parameter models.
Figure 4.3: Plots of $\log g_*(w)$ as in Example B for fixed $\eta = 0.8$ and: top) $\alpha = 3$ (note that $\alpha > 2\eta$ and that the graph is convex) and bottom) $\alpha = 1$ (note that $\alpha < 2\eta$ and that the graph is concave).
Example C: A joint tail model based on a modification of the mixed dependence structure.

As discussed in the previous example, boundary masses are presented here as they will be useful later, in Chapter 6, but for our joint tail models they play no role and can be then chosen arbitrarily.

Consider the measure density defined by

\[ h_\eta(w) = 2^{1/\eta} \eta^{-2}(\eta + 1) \]

for \( 0 < w < 1 \) and with atoms of equal mass \( 2^{1/\eta}(1 - \theta)/(\eta \theta) \) at \( w = 0 \) and \( w = 1 \), where \( \theta \in [0, 1] \) and \( \eta \in (0, 1] \). As in the previous example, the form of these boundary masses were chosen so that the results in Chapter 6 have a clearer relationship with their analogues for standard BEV distributions.

It is straightforward to verify that this density \( h_\eta \) satisfies the normalisation condition (4.10). Equation (4.7) then yields the model

\[ F_{ST}(s, t) = 2^{1/\eta}(s + t)^{-1/\eta} \]

for \( (s, t) \in [1, \infty) \times [1, \infty) \), and the associated limit function \( g_* \) is given by

\[ g_*(w) = 2^{1/\eta} \{w(1 - w)\}^{1/(2\eta)}. \]

This provides the following joint survivor model for \((X, Y)\):

\[ F_{XY}(x, y) = \lambda(2u)^{1/\eta} (x + y)^{-1/\eta} \]

for \((x, y) \in [u, \infty) \times [u, \infty) \), where \( u \) is a chosen high threshold and \( \lambda \) is the joint threshold exceedance probability. This model has only one dependence parameter \( \eta \) since its construction involves only the form of \( h_\eta \) in the interior of \([0, 1]\). The parameter \( \theta \) plays no role in satisfying condition (4.10).

Example D: Joint tail models based on discrete measures.

We start by considering the simple case where the measure has two atoms of equal mass \( a \) at positions \( w_1 \) and \( (1 - w_1) \) in the interior of \([0, 1]\), where \( 0 < w_1 < 1/2 \).

In order to satisfy the normalisation condition (4.10) the mass \( a \) has to be such that

\[ a = (2\eta)^{-1} w_1^{-1/\eta}. \]

Then, working through the construction of equation (4.7), we obtain

\[ F_{ST}(s, t) = \begin{cases} 
\frac{1}{2} (w_1 s)^{-1/\eta} \left\{ w_1^{1/\eta} + (1 - w_1)^{1/\eta} \right\} & \text{if } s/(s + t) < w_1, \\
\frac{1}{2} (s^{-1/\eta} + t^{-1/\eta}) & \text{if } s/(s + t) \in [w_1, 1 - w_1], \\
\frac{1}{2} (w_1 s)^{-1/\eta} \left\{ w_1^{1/\eta} + (1 - w_1)^{1/\eta} \right\} & \text{if } s/(s + t) > 1 - w_1.
\end{cases} \]
Chapter 4. A pseudo-polar representation of asymptotic independence

So, the limit function \( g_\ast \) has the form

\[
g_\ast(w) = \begin{cases} 
\frac{1}{2} \left\{ w(1-w) \right\}^{1/(2\eta)} \left\{ w_1(1-w) \right\}^{-1/\eta} \left\{ w_1^{1/\eta} + (1-w_1)^{1/\eta} \right\} & \text{if } w < w_1, \\
\frac{1}{2} \left\{ w(1-w) \right\}^{1/(2\eta)} \left\{ 1-w \right\}^{-1/\eta} \left\{ w_1^{1/\eta} + (1-w_1)^{1/\eta} \right\} & \text{if } w \in [w_1, 1-w_1], \\
\frac{1}{2} \left\{ w(1-w) \right\}^{1/(2\eta)} (w_1w)^{-1/\eta} \left\{ w_1^{1/\eta} + (1-w_1)^{1/\eta} \right\} & \text{if } w > 1-w_1,
\end{cases}
\]

and the corresponding parametric joint tail model for \((X, Y)\) is given by

\[
\bar{F}_{XY}(x, y) = \begin{cases} 
\frac{(\lambda w^{1/\eta}/2)}{2} \left\{ w_1^y \right\}^{-1/\eta} \left\{ w_1^{1/\eta} + (1-w_1)^{1/\eta} \right\} & \text{if } x/(x+y) < w_1, \\
\frac{(\lambda w^{1/\eta}/2)}{2} \left\{ x^{1/\eta} + y^{1/\eta} \right\} & \text{if } x/(x+y) \in [w_1, 1-w_1], \\
\frac{(\lambda w^{1/\eta}/2)}{2} \left\{ w_1y \right\}^{-1/\eta} \left\{ w_1^{1/\eta} + (1-w_1)^{1/\eta} \right\} & \text{if } x/(x+y) > 1-w_1,
\end{cases}
\]

where \( u \) is a chosen high threshold and \( \lambda \) is the joint threshold exceedance probability.

The above measure can be generalised to have \( 2n \) atoms of positive mass. Again, we focus on the symmetric case. Consider the discrete measure function that has positive masses \( a_1, a_2, ..., a_n \) at positions \( 0 < w_1 < w_2 < \cdots < w_n < 1/2 \) and also at positions \( 1-w_1, 1-w_2, ..., 1-w_n \), respectively. This measure satisfies the normalisation condition (4.10) if and only if

\[
\sum_{i=1}^{n} a_iw_i^{1/\eta} = (2\eta)^{-1}.
\]

More generally, considering the asymmetric discrete measure with positive masses \( a_1, ..., a_q, ..., a_n \) at positions \( 0 < w_1 < \cdots < w_q < \cdots < w_n < 1 \), where \( q \) is the first index such that \( w_i > 1/2 \) for all \( i \geq q \), we have that the normalisation condition (4.10) is satisfied if and only if

\[
\sum_{i=1}^{q-1} a_iw_i^{1/\eta} + \sum_{i=q}^{n} a_i(1-w_i)^{1/\eta} = \eta^{-1}.
\]

4.4 Marginal properties of \( \bar{F}_{ST} \)

In this section the marginal behaviour of the joint survivor function \( \bar{F}_{ST} \), defined in equation (4.5), is examined.

Although \( \bar{F}_{ST} \) is obtained from a distribution with unit Fréchet margins, its margins are not unit Fréchet distributed but are given by

\[
\Pr(S > s) = g(s, 1)s^{-1/(2\eta)} \quad \text{and} \quad \Pr(T > t) = g(1, t)t^{-1/(2\eta)}.
\]
Chapter 4. A pseudo-polar representation of asymptotic independence

To study the tail behaviour of these marginal survivor functions the structure of the ray dependence function $g$ is needed as $s \to \infty$ or $t \to \infty$ or equivalently the form of $g_s(w)$ is required as $w \to 1$ or $w \to 0$, respectively.

Defining the functions $I_1$ and $I_2$ by

$$I_1(z) = \int_0^z w^{1/\eta} h_\eta(w) \, dw \quad \text{and} \quad I_2(z) = \int_z^1 (1 - w)^{1/\eta} h_\eta(w) \, dw,$$

then $\Pr(S > s)$ and $\Pr(T > t)$ can be written as

$$\Pr(S > s) = g(s, 1)s^{-1/(2\eta)} = \eta \left[ s^{-1/\eta} I_1 \left( s/(s + 1) \right) + I_2 \left( s/(s + 1) \right) \right]$$

and

$$\Pr(T > t) = g(1, t)t^{-1/(2\eta)} = \eta \left[ I_1 \left( 1/(1 + t) \right) + t^{-1/\eta} I_2 \left( 1/(1 + t) \right) \right].$$

Thus, the tails of the variables $S$ and $T$ depend on the functions $I_1$ and $I_2$ and so depend on the particular form of $h_\eta$. Therefore, no general results are available and then the use of the previous examples is helpful here.

For both the modified logistic and asymmetric logistic models of Examples A and B, the joint survivor function $F_{ST}$ has margins satisfying

$$\Pr(S > s) = \begin{cases} O(s^{-1/\eta}) & \text{if } \alpha \leq \eta \\ O(s^{-1/\alpha}) & \text{if } \alpha > \eta \end{cases}$$

For the mixed joint survivor function $F_{ST}$ as in Example C, the margins satisfy $\Pr(S > s) = O(s^{-1/\eta})$. This shows that the marginal distributions can have tails heavier than the joint tail e.g. when $\alpha > \eta$.

Our examples suggest that margins behave like a power of $s$ (or $t$, respectively). However, marginal behaviour depends on the particular form of $h_\eta$ in general.

4.5 Standard bivariate extreme value (BEV) case

The procedure used above can also be applied to the standard BEV case. Consider the sequence of iid bivariate random variables $(X_1, Y_1), \ldots, (X_n, Y_n)$ with unit Fréchet margins and joint distribution function $F_{XY}$, where $F_{XY}$ is a BEV distribution so that $F_{XY}(x, y) = \exp \{-V(x, y)\}$ where $V(x, y) = \int_0^1 \max \left( \frac{w}{x}, \frac{1-w}{y} \right) h(w) \, dw$ for a non-negative measure density $h$ satisfying

$$\int_0^1 \omega h(w) \, dw = \int_0^1 (1 - w) h(w) \, dw = 1, \quad (4.22)$$

83
see Section 1.4.3, in Chapter 1.

Let \( u \) denote a high threshold, \( F \) represent the unit Fréchet distribution function and consider the behaviour of the limiting variable \((S,T)\) defined, as before, such that for all \((s,t) \in [1, \infty) \times [1, \infty)\)

\[
\Pr(S > s, T > t) = \lim_{u \to \infty} \frac{\Pr(X > su, Y > tu)}{\Pr(X > u, Y > u)} = \lim_{u \to \infty} \frac{1 - F(su) - F(tu) + F_{XY}(su, tu)}{1 - 2F(u) + F_{XY}(u, u)} = \frac{s^{-1} + t^{-1} - V(s,t)}{2 - V(1,1)},
\]

(4.23)
since \( V \) is homogeneous of order \(-1\). Re-arranging (4.23), we obtain

\[
\Pr(S > s, T > t) = (st)^{-1/2} \left( \frac{(s/t)^{1/2} + (i/s)^{1/2}}{2 - V(1,1)} \right) - \frac{g(s,t)}{(st)^{1/2}},
\]

(4.24)
which expresses \( \overline{F}_{ST} \) in terms of the components we had previously, on setting \( \eta = 1 \).

Transforming to the pseudo-radial and angular coordinates \( R = S + T \) and \( W = S/R \) and assuming that the density of \((R,W)\) exists, this density can be shown to satisfy \( f(r,w) = r^{-2}h^*(w) \) for \( w \in (0,1) \) and \( r \in [\max\{1/w, 1/(1-w)\}, \infty) \), where the function \( h^* \) is a non-negative measure density on \([0,1]\) determined by \( g \). Note that this density \( h^* \) and the previously defined density \( h \), satisfying condition (4.22), are not the same. Again, our aim is to express \( \Pr(S > s, T > t) \) from \( h^*(w) \) as our starting point.

Letting \( r^* = \max\{s/w, t/(1-w)\} \), we have

\[
\Pr(S > s, T > t) = \int_{w=0}^{1} \int_{r^*}^{\infty} r^{-2}h^*(w) \, dr \, dw = \int_{0}^{1} \min\left(\frac{w}{s}, \frac{1-w}{t}\right) h^*(w) \, dw = s^{-1} \int_{0}^{1} w h^*(w) \, dw + t^{-1} \int_{1/w}^{1} (1-w) h^*(w) \, dw.
\]

(4.25)
Thus, from a starting point \( h^*(w) \), we can use this representation, as before, to obtain parametric models for \( g \) or for the joint tail of the survivor function \( \overline{F}_{XY} \).

Writing \( s = t = t_0 \) in equation (4.25) and exploiting equation (4.24), we obtain

\[
1 = \int_{0}^{1/2} w h^*(w) \, dw + \int_{1/2}^{1} (1-w) h^*(w) \, dw.
\]

(4.26)
Chapter 4. A pseudo-polar representation of asymptotic independence

As in the asymptotically independent case, this equation yields a normalisation condition.

The joint tail of the original bivariate variable \((X, Y)\) may then be described by the model \(\overline{F}_{XY}(x, y) = \lambda \overline{F}_{ST}(x/u, y/u)\), for \(x > u\) and \(y > u\), where \(\lambda = \Pr(X > u, Y > u)\). This survivor model is defined in terms of the measure density \(h^*\) that has to satisfy the normalising condition (4.26). As seen above, equation (4.22) also gives a normalisation condition but in terms of the measure density \(h\). To compare these two normalisation conditions, the relationship between the densities \(h\) and \(h^*\) is examined next.

### 4.5.1 Relationship between the measure densities \(h\) and \(h^*\)

In order to compare the normalisation conditions given in (4.22) and (4.26) we derive here the relationship between the measure densities \(h\) and \(h^*\) defined above. In the following analysis we assume that both densities \(h\) and \(h^*\) exist, although more generally there may be situations where one exists and the other does not.

Noting that \(\overline{F}_{XY}\) is defined above as being a BEV distribution it can be derived that

\[
V(s, t) = \lim_{u \to \infty} \left\{ \frac{1}{s} + \frac{1}{t} - u \overline{F}_{XY}(us, ut) \right\}
\]

and, since \(\overline{F}_{XY}(u, u) \sim \frac{2 - V(1, 1)}{u}\), it follows that

\[
\lim_{u \to \infty} u \overline{F}_{XY}(us, ut) = (2 - V(1, 1)) \Pr(S > s, T > t),
\]

where \((S, T)\) is as in (4.23). Thus, we obtain

\[
\frac{1}{s} + \frac{1}{t} - V(s, t) = (2 - V(1, 1)) \Pr(S > s, T > t) \tag{4.27}
\]

for \(s > 1\) and \(t > 1\). Coles and Tawn (1991) showed, for the pseudo-polar coordinates \(r = s + t\) and \(w = s/r\), that

\[
\frac{\partial^2 V}{\partial s \partial t} = -r^{-3} h(w), \tag{4.28}
\]

so differentiating both sides of equation (4.27) and simplifying it we obtain

\[
h(w) = (2 - V(1, 1)) h^*(w). \tag{4.29}
\]

---

\(^{a}\)For example, in the case of independence \(dH\) places unit mass at the boundaries \(w = 0\) and \(w = 1\), while the density \(h^*\) exists.
Chapter 4. A pseudo-polar representation of asymptotic independence

Using this relationship between the two densities, the normalisation condition (4.26) can be now written in terms of \( h \) as follows

\[
\int_0^1 w h(w) \, dw + \int_0^1 (1 - w) h(w) \, dw = 2
\]

which simplifies to

\[
\int_0^1 h(w) \, dw = 2. \tag{4.30}
\]

It is clear that condition (4.22) implies this latter condition and that if \( h \) is symmetric then conditions (4.22) and (4.26) are equivalent. Thus, the measure densities associated with the logistic, asymmetric logistic and mixed models, as defined in Section 1.4.4, automatically satisfy condition (4.30).

4.6 Particular case: \( \eta = 1 \)

In this section we verify that the joint tail models obtained for the variables \((S,T)\) and \((X,Y)\) using the logistic, asymmetric logistic and mixed standard BEV dependence structure are identical respectively to those obtained from Examples A, B and C, defined in Section 4.3, by setting \( \eta = 1 \). Thus, the models given previously may be viewed as extensions of existing results and these BEV based models as particular special cases.

Examples:

A. Logistic dependence structure

The logistic BEV distribution is given by

\[
F_{XY}(x, y) = \exp \left\{ - \left( x^{-1/\alpha} + y^{-1/\alpha} \right)^\alpha \right\}. \tag{4.31}
\]

So following the steps given in Section 4.5 we first obtain

\[
\Pr(S > s, T > t) = \lim_{u \to \infty} \Pr(X > su, Y > tu) \Pr(X > u, Y > u)
\]

\[
= \frac{s^{-1} + t^{-1} - \left( s^{-1/\alpha} + t^{-1/\alpha} \right)^\alpha}{2 - 2^\alpha}.
\]

From here we exploit equation (4.11) and obtain

\[
F_{XY}(x, y) = \frac{\lambda u}{2 - 2^\alpha} \left\{ x^{-1} + y^{-1} - \left( x^{-1/\alpha} + y^{-1/\alpha} \right)^\alpha \right\},
\]

which is identical to equation (4.14) in Example A, on setting \( \eta = 1 \). Noting now that \( \lambda \sim \{2 - V(1,1)\} / u = (2 - 2^\alpha) / u \) this simplifies to

\[
F_{XY}(x, y) \sim x^{-1} + y^{-1} - \left( x^{-1/\alpha} + y^{-1/\alpha} \right)^\alpha.
\]
for large $x$ and $y$, which has exactly the same form as the standard BEV logistic joint survivor function, see e.g. Bruun and Tawn (1998).

B. Asymmetric logistic dependence structure

The asymmetric logistic BEV distribution is given by

$$F_{XY}(x,y) = \exp \left( \frac{1}{1-\alpha} \left[ \left( \frac{x}{\theta} \right)^{-1/\alpha} + \left( \frac{y}{\phi} \right)^{-1/\alpha} \right] \right).$$

Following the steps given in Section 4.5 and exploiting equation (4.11) we obtain

$$F_{XY}(x,y) = \frac{\lambda^u}{\theta + \phi - (\theta^{1/\alpha} + \phi^{1/\alpha})^\alpha} \left[ \left( \frac{x}{\theta} \right)^{-1} + \left( \frac{y}{\phi} \right)^{-1} - \left\{ \left( \frac{x}{\theta} \right)^{-1/\alpha} + \left( \frac{y}{\phi} \right)^{-1/\alpha} \right\}^\alpha \right],$$

which is identical to equation (4.17) (and then equivalent to model (4.19)) in Example B, on setting $\eta = 1$. Noting now that $\lambda \sim \theta + \phi - (\theta^{1/\alpha} + \phi^{1/\alpha})^\alpha / u$ this simplifies to

$$F_{XY}(x,y) \sim \left( \frac{x}{\theta} \right)^{-1} + \left( \frac{y}{\phi} \right)^{-1} - \left\{ \left( \frac{x}{\theta} \right)^{-1/\alpha} + \left( \frac{y}{\phi} \right)^{-1/\alpha} \right\}^\alpha$$

for large $x$ and $y$, which has exactly the same form as the standard BEV asymmetric logistic joint survivor function.

C. Mixed dependence structure

The mixed BEV distribution is given by

$$F_{XY}(x,y) = \exp \left[ - \left\{ x^{-1} + y^{-1} - \theta/(x+y) \right\} \right].$$

Following the steps given in Section 4.5, we obtain $F_{XY}(x,y) = 2u^\alpha (x+y)^{-1}$, which is identical to equation (4.21) in Example C, on setting $\eta = 1$, and since $\lambda \sim \theta/(2u)$ this simplifies to $F_{XY}(x,y) \sim \theta(x+y)^{-1}$, for large $x$ and $y$, the form of the standard BEV mixed joint survivor function.

These results show that our new models include joint survivor models based on the existing standard BEV case and also that the normalising constants present in the new models, namely $(2 - 2^{\alpha/\eta})^{-1}$, $N_{\theta\phi}^{-1}$ and $2^{1/\eta}g^{-1}$, appear naturally.

4.7 Conclusions

Ledford and Tawn (1997) showed the need for asymptotically independent models. Their models were the first to accommodate asymptotic dependence, asymptotic independence and negative association of marginal extreme values within one framework.
However, their parametric examples do not always provide valid joint densities. Our models are the first 'proper' fully parametric joint tail models that accommodate all these cases.

In contrast to the standard BEV approach, which concentrates on the distributional convergence of the normalised componentwise maxima, the framework used here is based on modelling joint tails and focuses directly on the tail structure of the joint survivor function. This yields very simple, tractable and easy to use parametric models with a completely characterised form in terms of an essentially arbitrary measure density $h_\eta$ that satisfies a simple normalising condition. Our findings also allow us to identify new classes of tail behaviour such as asymptotically convex or concave ray dependence.

A simple asymmetric parametric model was derived with asymmetry governed by a single parameter $\varrho$. This is important for inference (see Chapter 5). Furthermore, from the new parametric models full likelihood estimation is possible, i.e. joint estimation of marginal and dependence parameters, allowing exchange of information between margins.

In conclusion, significant extensions of both the theoretical and applicable tools of joint tail modelling have been obtained. Analogous point process theory will be developed in Chapter 6.
Chapter 5

Asymptotically independent joint tail modelling in practice

A fundamental issue in applied multivariate extreme values (MEV) analysis is modelling dependence within joint tail regions. In Chapter 4 we developed a pseudo-polar framework for modelling extremal dependence that extends the existing classical results and provides a constructional procedure for obtaining parametric joint tail dependence models. The practical application of such a model using the symmetric and asymmetric logistic dependence structures of Examples A and B is the focus of this chapter. We concentrate on the bivariate case again and cover applications to simulated and environmental data, detailing joint estimation of dependence and marginal parameters via likelihood methodology. Inference results are also developed in this chapter.

5.1 The modelling framework

First we develop the modelling framework assuming fixed unit Fréchet margins and then extend the set-up to unknown margins, the tails of which we model using the GPD.

5.1.1 Known unit Fréchet margins

Suppose that an iid sequence of bivariate random variables \( \{(X_i, Y_i); i = 1, \ldots, n\} \) is given from the joint distribution function \( F_{XY} \) with unit Fréchet margins and unknown
dependence structure. The goal here is to estimate the joint tail of $F_{XY}$, or equivalently to model the extremal dependence structure, since the margins are known.

Following the approach of Chapter 4, the joint survivor function $\overline{F}_{XY}$ is assumed to satisfy

$$\overline{F}_{XY}(x, y) = \mathcal{L}(x, y)(xy)^{-1/(2\eta)}.$$  

Since the marginal variables are standardised, the parameter $\eta$ provides a measure of the dependence between the marginal tails. In particular, if $1/2 < \eta \leq 1$ then the marginal variables are positively associated; they are independent when $\eta = 1/2$; and if $0 < \eta < 1/2$ then the variables are negatively associated.

Next, in order to estimate the joint tail of $F_{XY}$, a high threshold $u$ needs to be chosen. Note that the same threshold $u$ is considered for the two margins since they are both unit Fréchet distributed. Then, treating the limit in equation (4.3) as an approximation in the joint tail, the bivariate random variables $(S_i, T_i) = (X_i/u, Y_i/u)(X_i > u, Y_i > u)$ are constructed. We concentrate first on the parametric modified logistic model from Example A in Chapter 4, that is, we work with the joint survivor function $\overline{F}_{ST}$ with the form

$$\overline{F}_{ST}(s, t) = \frac{1}{2} - \frac{1}{2^2} \left\{ s^{-1/\eta} + t^{-1/\eta} - \left( s^{-1/\alpha} + t^{-1/\alpha} \right)^{\alpha/\eta} \right\} \tag{5.1}$$

for $s, t > 1, \eta \in (0, 1]$ and $\alpha > 0$. This provides the following parametric model for the joint tail of the survivor function of the original variable $(X, Y)$

$$\overline{F}_{XY}(x, y) = \frac{\lambda u^{1/\eta}}{2} \left\{ x^{-1/\eta} + y^{-1/\eta} - \left( x^{-1/\alpha} + y^{-1/\alpha} \right)^{\alpha/\eta} \right\} \tag{5.2}$$

for $x, y > u$, where $\lambda$ is the joint threshold exceedance probability.

When data exhibit some asymmetry in a joint tail region, i.e. when the underlying distribution function has an asymmetric joint tail, it is more appropriate to use an asymmetric parametric model for the statistical modelling. Accordingly, we use the modified asymmetric logistic model in cases where there is evidence of asymmetry, and thus work with joint survivor function for the variable $(S, T)$ of the form

$$\overline{F}_{ST}(s, t) = N_\theta^{-1} \left[ s^{-1/\eta} + \left( \frac{t}{\theta} \right)^{-1/\eta} - \left\{ s^{-1/\alpha} + \left( \frac{t}{\theta} \right)^{-1/\alpha} \right\}^{\alpha/\eta} \right] \tag{5.3}$$

for $s, t > 1, \eta \in (0, 1], \alpha, \theta > 0$, where $N_\theta = 1 + \theta^{1/\eta} - (1 + \theta^{1/\alpha})^{\alpha/\eta}$. Similar to the above case, the joint tail of $\overline{F}_{XY}$ is modelled by

$$\overline{F}_{XY}(x, y) = \frac{\lambda u^{1/\eta}}{N_\theta} \left[ x^{-1/\eta} + \left( \frac{y}{\theta} \right)^{-1/\eta} - \left\{ x^{-1/\alpha} + \left( \frac{y}{\theta} \right)^{-1/\alpha} \right\}^{\alpha/\eta} \right] \tag{5.4}$$
Chapter 5. Asymptotically independent joint tail modelling in practice

for \( x, y > u \), where \( \lambda \) is as in the previous case.

The parametric models described above have respectively two and three dependence parameters. Broadly interpreted, the dependence parameters \( \eta, \alpha \) and \( \varphi \) correspond to measures of joint tail heaviness, ray dependence and asymmetry, respectively.

5.1.2 Unknown margins

So far, attention has been restricted to modelling dependence between the tails of variables which are known to be unit Fréchet distributed. In applications to real data the marginal distributions are unknown and also need to be estimated. In this section, we demonstrate how standard univariate tail models and the dependence models given above can be combined in a fully parametric joint tail model.

Let \((X^*, Y^*), \ldots, (X^*_n, Y^*_n)\) be an iid sequence of bivariate random variables from the joint distribution function \( F_{X^*Y^*} \) with unknown marginal distribution functions \( F_1 \) and \( F_2 \). We assume that the marginal variables \( X^* \) and \( Y^* \) follow a generalised Pareto distribution (GPD) for \( X^* > u_1 \) and \( Y^* > u_2 \) with parameters \((\xi_1, \sigma_1, \lambda_1)\) and \((\xi_2, \sigma_2, \lambda_2)\), respectively, that is,

\[
F_1(x^*) = 1 - \lambda_1 \left\{ \frac{1 + \xi_1 (x^* - u_1)}{\sigma_1} \right\}^{1/\xi_1}
\]

and

\[
F_2(y^*) = 1 - \lambda_2 \left\{ \frac{1 + \xi_2 (y^* - u_2)}{\sigma_2} \right\}^{1/\xi_2}
\]  

(5.5)

for \( x^* > u_1 \) and \( y^* > u_2 \), where \( u_1 \) and \( u_2 \) are high marginal thresholds, \( \xi_i \) and \( \sigma_i > 0 \), for \( i = 1, 2 \), are respectively shape and scale parameters, and \( \lambda_i \) \((i = 1, 2)\) are marginal threshold exceedance probabilities.

It is easy to see that the variables \( X = -1/\log F_1(X^*) \) and \( Y = -1/\log F_2(Y^*) \) are unit Fréchet distributed. Thus, choosing \( u_1 \) and \( u_2 \) such that both empirical marginal threshold exceedance probabilities are equal and using these empirical probabilities in the definition of the distribution functions \( F_1 \) and \( F_2 \) in equations (5.5), we define \( u = -1/\log F_1(u_1) = -1/\log F_2(u_2) \) as the single threshold for both unit Fréchet variables \( X \) and \( Y \). The joint tail of the distribution of the variable \((X^*, Y^*)\) can be then approximated by the following full joint tail model

\[
F_{X^*Y^*}(x^*, y^*) = \lambda^{\log ST} \left\{ \frac{-1}{u \log F_1(x^*)}, \frac{-1}{u \log F_2(y^*)} \right\}
\]
for \(x^* > u_1\) and \(y^* > u_2\). Hence, using the GPD distribution functions \(F_1\) and \(F_2\) as in equations (5.5), the full joint tail model has the form
\[
\bar{F}_{X,Y}(x^*,y^*) = \frac{\lambda}{2^{1-2\alpha/\eta}} \left( \frac{-1}{u \log F_1(x^*)} \right)^{-1/\eta} \left( \frac{-1}{u \log F_2(y^*)} \right)^{-1/\eta} - \\
\left[ \left( \frac{-1}{u \log F_1(x^*)} \right)^{-1/\alpha} + \left( \frac{-1}{u \log F_2(y^*)} \right)^{-1/\alpha} \right]^{\alpha/\eta}
\]
(5.6)
for \(\bar{F}_{ST}\) as in equation (5.1), and
\[
\bar{F}_{X,Y}(x^*,y^*) = \frac{\lambda}{N} \left( \frac{-1}{u \log F_1(x^*)} \right)^{-1/\eta} \left( \frac{-1}{u \log F_2(y^*)} \right)^{-1/\eta} - \\
\left[ \left( \frac{-1}{u \log F_1(x^*)} \right)^{-1/\alpha} + \left( \frac{-1}{u \log F_2(y^*)} \right)^{-1/\alpha} \right]^{\alpha/\eta}
\]
(5.7)
for \(\bar{F}_{ST}\) as in equation (5.3). These fully parametric joint tail models have now nine and ten parameters, respectively. These comprise the GPD parameters \((\xi_i, \sigma_i, \lambda_i)\) for \(i = 1, 2\), the joint threshold exceedance probability \(\lambda\) and the dependence parameters \(\eta, \alpha\) and for the case of the asymmetric logistic model, \(\rho\).

In the next section we examine how these parameters may be estimated using maximum likelihood.

### 5.2 Likelihood

When fitting the joint tail model to data, both dependence and marginal parameters may need to be estimated. The threshold censored likelihood approach described in Section 1.5.3.3 may be used for this purpose.

#### 5.2.1 Known unit Fréchet margins

We start with the case where the margins are known to be unit Fréchet distributed and then consider data as in Section 5.1.1. Let us first divide the outcome space into the four regions
\[
\{R_{ij} : i = I(x \geq u), j = I(y \geq u)\}
\]
where \(I\) is the indicator function. Marginal observations that do not exceed the threshold \(u\) are considered censored at \(u\), and hence no assumptions are made about the dependence structure form outside \(R_{11}\).
Using the logistic survivor model as in equation (5.2), the likelihood contribution \( L_{ij}(x, y) \) of a point \((x, y)\) which falls into region \( R_{ij} \) is given by

\[
L_{00}(x, y) = 2\exp(-1/\alpha) - 1 + \lambda
\]

\[
L_{10}(x, y) = f(x) - \frac{\lambda u_1/\eta}{\eta(2 - 2\alpha/\eta)} \left\{ x^{-(1+1/\eta)} - \left( x^{-1/\alpha} + u_1^{-1/\alpha} \right)^{\alpha/\eta - 1} x^{-(1+1/\alpha)} \right\}
\]

\[
L_{01}(x, y) = f(y) - \frac{\lambda u_1/\eta}{\eta(2 - 2\alpha/\eta)} \left\{ y^{-(1+1/\eta)} - \left( u_1^{-1/\alpha} + y^{-1/\alpha} \right)^{\alpha/\eta - 1} y^{-(1+1/\alpha)} \right\}
\]

\[
L_{11}(x, y) = \frac{\lambda u_1/\eta(\eta - \alpha)}{\alpha\eta^2(2 - 2\alpha/\eta)} \left( x^{-1/\alpha} + y^{-1/\alpha} \right)^{\alpha/\eta - 2} \left( xy \right)^{-(1+1/\alpha)}
\]

where \( f \) represents the unit Fréchet density function. The overall likelihood \( L_n(\alpha, \eta) \) for a set of \( n \) points is then the product of the corresponding \( L_{ij} \) terms.

For the asymmetric logistic survivor model in equation (5.4) the corresponding likelihood contribution of a point \((x, y)\) in region \( R_{ij} \) is

\[
L_{00}(x, y) = 2\exp(-1/\alpha) - 1 + \lambda
\]

\[
L_{10}(x, y) = f(x) - \frac{\lambda u_1/\eta}{\eta(2 - 2\alpha/\eta)} \left\{ x^{-(1+1/\eta)} - \left( x^{-1/\alpha} + \frac{u_1}{\theta} \right)^{1/\alpha} x^{-(1+1/\alpha)} \right\}
\]

\[
L_{01}(x, y) = f(y) - \frac{\lambda u_1/\eta}{\eta(2 - 2\alpha/\eta)} \left\{ y^{-(1+1/\eta)} - \left( u_1^{-1/\alpha} + \frac{y}{\theta} \right)^{-1/\alpha} \left( \frac{y}{\theta} \right)^{-(1+1/\alpha)} \right\}
\]

\[
L_{11}(x, y) = \frac{\lambda u_1/\eta(\eta - \alpha)}{\alpha\eta^2(2 - 2\alpha/\eta)} \left( x^{-1/\alpha} + \frac{y}{\theta} \right)^{-1/\alpha - 1} \left( x \left( \frac{y}{\theta} \right) \right)^{-(1+1/\alpha)}
\]

where, again, \( f \) represents the unit Fréchet density function. Again, the overall likelihood \( L_n(\alpha, \eta, \Theta) \) for a set of \( n \) points is the product of the corresponding \( L_{ij} \) terms.

### 5.2.2 Unknown margins

When the marginal distributions are unknown we use the GPD to model their tails. Considering data as in Section 5.1.2 and defining the regions

\[
\{ R_{ij}^* : i = I(x^* \geq u_1), j = I(y^* \geq u_2) \},
\]

the fully parametric joint survivor models defined in equations (5.6) and (5.7) are assumed to hold exactly in the joint tail region \( R_{11}^* \). For the logistic survivor model in equation (5.6), the likelihood contribution of a point \((x^*, y^*)\) which falls into region \( R_{ij}^* \)
Chapter 5. Asymptotically independent joint tail modelling in practice

is given by

\[ L_{00}(x^*, y^*) = 2 \exp(-1/n) - 1 + \lambda \]

\[ L_{10}(x^*, y^*) = f_1(x^*) - \frac{\lambda f_1(x^*)}{\eta(2 - 2^{\alpha/\eta})F_1(x^*)} \left( \left\{ \frac{-1}{\log F_1(x^*)} \right\}^{\alpha/\eta - 1} \right)^{-1} \]

\[ L_{01}(x^*, y^*) = f_2(y^*) - \frac{\lambda f_2(y^*)}{\eta(2 - 2^{\alpha/\eta})F_2(y^*)} \left( \left\{ \frac{-1}{\log F_2(y^*)} \right\}^{\alpha/\eta - 1} \right)^{-1} \]

\[ L_{11}(x^*, y^*) = \frac{\lambda U^{1/\eta} f_1(x^*) f_2(y^*)(\eta - \alpha)}{\alpha\eta^2(2 - 2^{\alpha/\eta})F_1(x^*)F_2(y^*)} \left( \left\{ \frac{-1}{\log F_1(x^*)} \right\}^{\alpha/\eta - 1} \left\{ \frac{-1}{\log F_2(y^*)} \right\}^{\alpha/\eta - 1} \right)^{-1} \]

where \( F_1 \) and \( F_2 \) are the GPD distribution functions defined in equations (5.5) and \( f_1 \) and \( f_2 \) are their respective density functions. The overall likelihood \( L_n(\alpha, \eta, \xi_1, \sigma_1, \xi_2, \sigma_2) \) for a set of \( n \) points is then the product of the corresponding \( L_{ij} \) terms.

The likelihood contributions in each region \( R_{ij}^* \) for the asymmetric logistic survivor model are obtained similarly. According to Smith (1985) there is no non-regularity in the maximum likelihood attributed to the GPD estimation when both shape parameters \( \xi_1 \) and \( \xi_2 \) are greater than \(-1/2\). This condition is assumed here, since the situation \( \xi_1 < -1/2 \) or \( \xi_2 < -1/2 \) does not prove to be very common in practice. More generally, when \(-1 < \xi_i < -1/2\) for \( i = 1, 2 \) then maximum likelihood estimates of the marginal parameters exist but are non-regular, whereas when \( \xi_i < -1 \) then maximum likelihood estimates may not even exist.

Note that points in region \( R_{11}^* \) contribute directly to both the generalised Pareto marginal distributions and the joint dependence model, whereas points in regions \( R_{10}^* \) and \( R_{01}^* \) contribute to the respective GPD and give only a small contribution to the joint dependence model. The only information conveyed by points in \( R_{00}^* \) is that they occur below the thresholds.

To ease the computational burden of maximising the above likelihoods, the joint and marginal threshold exceedance probabilities \( \lambda, \lambda_1 \) and \( \lambda_2 \) may be replaced by their empirical analogues at the outset. These values are held fixed when maximising the joint likelihood function over the other parameters.
5.3 Inference and diagnostics

The parametric models described in Section 4.3 can be exploited to develop likelihood ratio tests for asymptotic independence, ray independence and asymmetry. Such tests provide an aid to model selection, as they inform whether the use of an asymptotically independent joint tail model or an asymmetric joint tail model is merited.

5.3.1 Test for asymptotic independence

A natural framework for testing asymptotic dependence against asymptotic independence is provided by our parametric models. Let $L_n(\alpha, \eta)$ be the maximum of the likelihood obtained from equations (5.8) taken over the dependence parameters $\alpha > 0$ and $\eta \in (0, 1]$ and write $L_n(\alpha, 1)$ for the corresponding maximised likelihood under the constraint $\eta = 1$. Then, under the condition $\eta = 1$, we have

$$2 \log \{L_n(\alpha, \eta)/L_n(\alpha, 1)\} \overset{\mathbb{P}}{\rightarrow} Z^2 \quad \text{as } n \to \infty$$

where the non-negative random variable $Z$ has law

$$\Pr(Z \leq z) = h^*(z) \Phi(z)$$

for $h^*(\cdot)$ the Heaviside step function and $\Phi(\cdot)$ the standard normal distribution function (see Self and Liang, 1987 for the proof).

Consequently, for a given sample and size, asymptotic dependence will be rejected if $2 \log \{L_n(\alpha, \eta)/L_n(\alpha, 1)\} > c$ where $c$ is an appropriate critical value. For example, the critical value $c$ has values 2.7 and 5.4 for tests of sizes 0.05 and 0.01, respectively.

An analogous argument shows that result (5.11) is still valid for the full parameter likelihood $L_n(\alpha, \eta, \xi_1, \xi_2, \delta_1, \delta_2)$ obtained from equations (5.10) provided that the marginal regularity conditions of the maximum likelihood, as mentioned above, are satisfied. Similarly, a test for 'near' extremal independence, i.e. $\eta = 1/2$, is also possible and since $\eta = 1/2$ is an interior point of the parameter space $(0, 1]$, standard likelihood ratio tests apply.

5.3.2 Test for ray independence

The BSV function $L$ was termed asymptotically ray independent when the dependence function $g_*(w)$ is constant over different rays. For our modified logistic model, asym-
Chapter 5. Asymptotically independent joint tail modelling in practice

totic ray independence corresponds to $\alpha = 2\eta$. Therefore a test for ray independence against ray dependence can be obtained in the following way. Denote the maximum of the likelihood obtained from equations (5.8) under the restriction $\alpha = 2\eta$ by $L_n(\hat{\alpha}, \hat{\eta})$ and let $L_n(\hat{\alpha}, \hat{\eta})$ be the unrestricted maximised likelihood. Thus, under $\alpha = 2\eta$, we have

$$2 \log \left( \frac{L_n(\hat{\alpha}, \hat{\eta})}{L_n(\hat{\alpha}, \hat{\eta})} \right) \xrightarrow{\text{w}} \chi^2_{(1)}$$

as $n \to \infty$, (5.13)

where $\chi^2_{(1)}$ represents the chi-squared distribution with one degree of freedom (see e.g. Chernoff, 1954 or Self and Liang, 1987 for the proof).

A one-tail test for ray independence against convex ray dependence (or concave ray dependence) can also be established. Since asymptotic ray independence corresponds to $\alpha = 2\eta$ and asymptotic convex ray dependence to $\alpha > 2\eta$, we define $\delta \geq 0$ such that $\alpha = 2\eta + \delta$. Therefore, testing for ray independence against convex ray dependence corresponds to testing $\delta = 0$ against $\delta > 0$. Denoting the maximum of the likelihood obtained from equations (5.8) under the restriction $\delta = 0$ by $L_n(0)$ and letting $L_n(\hat{\delta})$ be the maximised likelihood under $\delta \geq 0$, we have, under $\delta = 0$, that

$$2 \log \left( \frac{L_n(0)}{L_n(\hat{\delta})} \right) \xrightarrow{\text{w}} Z^2$$

as $n \to \infty$, (5.14)

where $Z$ is as defined in equation (5.12). Similarly, results (5.13) and (5.14) are still valid for the full parameter likelihood $L_n(\hat{\alpha}, \hat{\eta}, \hat{\xi}_1, \hat{\xi}_2, \hat{\delta}_1, \hat{\delta}_2)$ obtained from equations (5.10) providing that the marginal regularity conditions of the maximum likelihood are satisfied.

5.3.3 Test for asymmetry

It was seen in Chapter 4 that the asymmetry of the modified asymmetric logistic model described in Example B is governed only by the parameter $\varphi > 0$ and that symmetry arises when $\varphi = 1$. Thus, from this asymmetric model, a test for asymmetry against symmetry can be derived. Let $L_n(\hat{\varphi}, \hat{\eta}, 1)$ denote the maximum of the likelihood obtained from equations (5.9) under the constraint $\varphi = 1$ and let $L_n(\hat{\varphi}, \hat{\eta}, \hat{\delta})$ denote the maximum of the likelihood taken over the dependence parameter $\varphi > 0$, thus

$$2 \log \left( \frac{L_n(\hat{\varphi}, \hat{\eta}, 1)}{L_n(\hat{\varphi}, \hat{\eta}, \hat{\delta})} \right) \xrightarrow{\text{w}} \chi^2_{(1)}$$

as $n \to \infty$, (5.15)

Similarly, result (5.15) also holds for the full likelihood function $L_n(\hat{\varphi}, \hat{\eta}, \hat{\delta}, \hat{\xi}_1, \hat{\xi}_2, \hat{\delta}_1, \hat{\delta}_2)$ from the full modified asymmetric logistic model providing that the regularity conditions of the maximum likelihood are satisfied. Note that it only makes sense, both
practically and mathematically, to test for asymmetry when $\alpha \neq 2\eta$, since otherwise the survivor model necessarily has a symmetric form. This is clear from the fact that ray independence implies symmetry, although the converse is not true.

### 5.3.4 Diagnostics

Fitting the model requires estimation of the dependence parameter $\eta$ and of course the other parameters. Here, we consider a diagnostic developed by Ledford and Tawn (1996) that enables $\eta$ to be estimated without carrying out the full model estimation. This diagnostic stage is useful as a preliminary step in exploring observed data, and also provides a useful starting point for numerical estimation of the full likelihood. Similarly, we also exploit the diagnostic for the limit function $g_*$ developed by Ledford and Tawn (1997) which allows the ray dependence form to be examined prior to the full analysis.

Define the univariate structure variable $V = \min(X, Y)$. By equation (4.1), it follows that $\Pr(V > u) = F_{XY}(u, u) = \mathcal{L}(u, u)u^{-1/\eta}$. Now, $\mathcal{L}$ is a slowly varying function so may be approximated by a constant $K$ and then it follows that $V$ approximately satisfies $\Pr(V > u) \sim K u^{-1/\eta}$ for large values of $u$. Standard univariate threshold based methods (see Section 1.3.5) can now be used to estimate $\eta$ as the shape parameter of the $V$-variable, see Ledford and Tawn (1996).

The diagnostic developed by Ledford and Tawn (1997) for the limit function $g_*$ is as follows. Assuming that the joint tail of $F_{XY}$ satisfies equation (4.1) and taking $c_1 = c_2 = 1/(2\eta)$, we have, for large $u$

$$
g_*(w) = \begin{cases} 
(1-w)/(2\eta) F_{XY}(u, (1-w)u/w) / F_{XY}(u, u) & \text{for } w \in (0, 1/2], \\
(1-w)/(2\eta) F_{XY}((1-w)u, w) / F_{XY}(u, u) & \text{for } w \in (1/2, 1).
\end{cases}
$$

(5.16)

Hence using the chosen threshold $u$, the estimate of $\eta$ obtained above and replacing $F_{XY}$ in equation (5.16) with empirical counts, the ray dependence function $g_*$ may be estimated. Approximate pointwise confidence intervals may be constructed by the delta-method.
5.4 Application to simulated data

In this section we apply our models to simulated data and then analyse their performance. We consider four data sets, each of 25,000 points \( \{(X_i, Y_i); i = 1, \ldots, 25,000\} \) unit Fréchet marginally distributed and with dependence structures as follows:

A. Bivariate extreme value (BEV) logistic dependence structure with \( \alpha = 0.75 \)

B. Bivariate normal dependence structure with correlation \( \rho = 0.5 \)

C. Morgenstern dependence structure with \( \beta = 0.75 \)

D. Bivariate normal dependence structure with correlation \( \rho = -0.5 \).

Ledford and Tawn (1997) showed that the coefficient of tail dependence \( \eta \) has values 1, 0.75, 0.5 and 0.25 for models A, B, C and D, respectively. In fact, componentwise maxima of data from model A are asymptotically dependent while extremes from the other three models are asymptotically independent, model B presenting positive association, model C being in the near independence case and model D exhibiting negative association. This can be confirmed by observing the logarithm scale scatter plots of the data sets in Figure 5.1.

Ledford and Tawn (1997) showed that all these examples are asymptotically ray independent except the BEV case which either exhibits ray dependence or has ray dependence function \( g^a(w) \) identical to 1 when margins are independent. For model A, the parameter \( \alpha \) has the value 0.75 while for the other three cases, \( \alpha \) within our tail model may perhaps be expected to equal approximately \( 2^{\hat{a}} \) since the joint tails of the underlying distributions for models B, C and D are asymptotically ray independent.

For the joint modelling, the first task is to choose an appropriate marginal threshold \( u \). Following the Ledford and Tawn (1996) approach, an appropriate value for \( u \) may be chosen constructing \( V_i = \min(X_i, Y_i) \) for each data set and selecting \( u \) to be the empirical 95% point of the \( V \)-variable. This choice was made by estimating the (shape)

---

1 See equation (3.2).

2 The Morgenstern distribution function is given by

\[
F(x, y) = F(x)F(y) \left(1 + \beta F(x)F(y) \right)
\]

where \(-1 < \beta \leq 1\) and \(F\) and \(F\) denote the unit Fréchet distribution and survivor functions respectively.
Figure 5.1: Plot of the four sets of simulated data with the selected thresholds included (on a logarithmic scale): a) data set A (BEV logistic, $\alpha = 0.75$); b) data set B (bivariate normal, $\rho = 0.5$); c) data set C (Morgenstern, $\beta = 0.75$) and d) data set D (bivariate normal, $\rho = -0.5$).
parameter $\eta$ from the univariate sample $V_1, \ldots, V_n$ and seeking the threshold above which those estimated values are stable. This threshold selection has the advantage that model fitting is based on the same number of joint threshold exceedances in each case, 1250, which provides similar precision of calculated quantities from the fitted models. The values of the threshold $u$ are then 7.42, 6.93, 5.03 and 2.43 for data sets A, B, C and D, respectively, and they are included on the scatter plots in Figure 5.1.

The results of our data analyses are presented next. First, diagnostic results are considered and then results obtained by fitting our joint tail model are shown.

5.4.1 Inference and results

The likelihood ratio statistic defined in result (5.11) for testing asymptotic dependence versus asymptotic independence has the values 0.3, 93.6, 418 and 1071.4 for data sets A, B, C and D, respectively. Comparing with both the 95% and the 99% critical values, 2.7 and 5.4, only the value for data set A is non-significant, and the values for all other sets are clearly significant. These findings are as expected and confirm that the joint tails of B, C and D exhibit asymptotic independence.

The likelihood ratio statistic defined in result (5.15) for testing asymmetry has values 1.21, 0.64, 0.08 and 0.32 for A, B, C and D, respectively. All of these values are non-significant, as expected, since the underlying distributions are symmetric. Accordingly, we will use here the symmetric logistic joint tail model defined in equation (5.2).

Ray independence can also be tested using the likelihood ratio statistic in result (5.13), which has values 662.8, 130.0, 0.06 and 0.36 for A, B, C and D, respectively. We thus conclude that data sets A and B have ray dependence. This is as expected for A since the BEV joint tail is ray dependent. However, for data set B, although the asymptotic model is ray independent, the tail exhibits ray dependence at the sub-asymptotic threshold used here. For the remaining cases C and D, the values of the likelihood ratio statistic are non-significant indicating no significant ray dependence, as expected.

Parameter estimates obtained by maximising the likelihood function defined in equations (5.8) for the chosen thresholds and corresponding standard errors based on the delta-method (in parentheses) are given in Table 5.1. As discussed above, the empirical joint threshold exceedance probability was used to estimate $\lambda$. Diagnostic-based
estimates of $\eta$ obtained by fitting a GPD model to the univariate $V$-variable are also included in Table 5.1.

<table>
<thead>
<tr>
<th>Data set</th>
<th>$\eta$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
<td>Diag.</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>1.00 (0.06)</td>
</tr>
<tr>
<td>B</td>
<td>0.75</td>
<td>0.77 (0.05)</td>
</tr>
<tr>
<td>C</td>
<td>0.5</td>
<td>0.48 (0.04)</td>
</tr>
<tr>
<td>D</td>
<td>0.25</td>
<td>0.27 (0.03)</td>
</tr>
</tbody>
</table>

Table 5.1: Dependence parameter estimates obtained using the chosen threshold $u$ for each data set. For $\eta$, true values and estimates obtained from the diagnostic and model fits are given whereas for $\alpha$ the estimates are obtained from the model fit. Standard errors are given in parentheses.

For data sets A, B and C model and diagnostic estimates of $\eta$ are equally good. However, for data set D, the model estimate for $\eta$ is not as good as the diagnostic estimate, which may be caused by the threshold used being low. Standard errors are smaller for estimates obtained by the model fit, as expected. The estimate for the dependence parameter $\alpha$ has small bias for data set A. For data set C, the estimate of $\alpha$ is close to $2\eta$ whereas for data set D, although the estimate of $\alpha$ is not twice the true value of $\eta$, it is twice the value of the model estimate of $\eta$. Better results can be achieved using higher thresholds as can be seen from Figure 5.2 which contains pointwise estimates of the dependence parameters $\eta$ and $\alpha$ calculated using the new model for a range of $V$-variable threshold probabilities for each data set. For comparison with the model estimates for $\eta$, this figure also includes the diagnostic estimates of $\eta$ for the chosen range of thresholds. All estimates are given together with 95% confidence intervals based on their standard errors.

Figure 5.2 also allows us to verify the adequacy of the fitted model as well as the chosen threshold, showing some stability in the model based parameter estimates for thresholds above the chosen structure variable threshold probability 0.95. Apart from those for the BBV data, estimates of $\eta$ obtained from the model fit tend to be bigger than the
Chapter 5. Asymptotically independent joint tail modelling in practice

Figure 5.2: Model based estimates (solid lines) of the dependence parameters $\eta$ and $\alpha$ together with diagnostic estimates of $\eta$ (long dashed) for a range of probability thresholds of the variable $V = \min(X, Y)$ for a) data set A; b) data set B; c) data set C and d) data set D. In this figure 95% confidence intervals for the model based estimates (dotted) and for the diagnostic estimates (short dashed) are included. The true values of $\eta$ and the true value of $\alpha$ for a) are represented by a straight line.
true values. Comparing these estimates of \( \eta \) to those obtained by the diagnostic we conclude that model and diagnostic estimates are equally good for data sets A and B over the whole range of thresholds used and for data C when using high thresholds. Figure 5.2 also shows that the diagnostic produces better estimates of \( \eta \) than the model fit for data set D and when using low thresholds for data set C. It is also clear from the figure that confidence intervals for the model based estimates are narrower than those for the diagnostic estimates. Plots of \( \alpha \) estimates against threshold exhibit small bias for data set A and values close to \( 2\eta \) for data set C. For the remaining data sets, estimates of \( \alpha \) have values smaller than \( 2\eta \) for data set B and bigger than \( 2\eta \) for data set D, although in this latter case the values are twice the value of the model estimates of \( \eta \).

The adequacy of the model can also be investigated by the limit function \( g_* \) which assesses the form of ray dependence in \( R_{11} \). Figure 5.3 depicts pointwise estimates of \( g_*(w) \) obtained from the non-parametric estimator (5.16) using empirical counts, together with the fitted \( g_*(w) \)-function for our model for each data set. Approximate 95% confidence intervals for the diagnostic estimates of \( g_* \) constructed by the delta-method are included for data sets C and D only, as very narrow confidence intervals are obtained for the other data sets.

The fitted ray dependence function agrees closely with the observed \( g_* \) given by the diagnostic, capturing the positive association for both A and B. However, the fitted \( g_* \) does not capture completely the observed convexity for C and D. In fact, although the underlying distributions for C and D are asymptotically ray independent, i.e. \( g_*(w) = 1 \), C shows some residual ray variation and D exhibits pronounced convex ray dependence, suggesting that the chosen thresholds are considerably sub-asymptotic. The discrepancy between the fitted and diagnostic estimates of \( g_* \) is not very significant for C if we take into account that the fitted \( g_* \) lies within the confidence interval. For D, the diagnostic estimate of \( g_* \) is strongly ray dependent while the model estimate is more consistent with the ray independence of the underlying distribution. This inconsistency of the observed and fitted ray dependence model is reflected in the model estimates of \( \eta \) given in Table 5.1, and indeed, the model estimate of \( \eta \) for D is poorer than the diagnostic one. This can be explained by the ray dependence parametric model not being flexible enough to capture the observed convexity, and thus resulting in bias. Using our model at higher thresholds improves the \( g_* \) estimation difficulties identified above,
Figure 5.3: Plots of the fitted function $g_*(w)$ for the new model (dashed), together with the diagnostic estimate of $g_*$ obtained using empirical counts (solid line) for a) data set A; b) data set B; c) data set C; d) data set D. Approximate 95\% delta-method based confidence intervals (dotted) are included for data sets C and D only, as very narrow confidence intervals are obtained for the other data sets.
providing there is enough data to do so.

Further examination of the performance of the new model is carried out by constructing the joint probability density function of the fitted model within region $R_{11}$ by differentiation. Contour plots of this joint density for each data set are shown in Figure 5.4 using a logarithmic scale (i.e. Gumbel margins), and for contour levels at $10^{-j/2}$ for $j = 4, \ldots, 11$. The correspondence between the joint density and the observed frequency of points from each data set is clear.

### 5.4.2 Extrapolation

As was mentioned in Chapter 2, estimation of extreme quantiles is often the main requirement of an extreme analysis and therefore is frequently the key factor for judging the performance of the model used. Thus, in this section, we show how the fitted model extrapolates from the sample information, constructing the joint tail contour curves of the survivor function from the fitted model for each data set. That is, for a given small joint tail probability $p$, we plot the contour $F_{XY}(x, y) = p$, where $F_{XY}$ is the survivor function defined in equation (5.2). These curves together with contours of the true joint tail survivor function are plotted in Figure 5.5 on a logarithmic scale, for each data set. The associated data points lying in region $(u, \infty) \times (v, \infty)$ are also included for each case in order to indicate how the fitted model extrapolates beyond the range of the data to which the model has been fitted. Contours were plotted at levels $p = 10^{-i}$ for $i = 2, \ldots, 7$.

For data sets A, B and C our model performs well, giving estimates that are almost coincident with the true curves for data set A. For D the contour curves of the fitted model slightly overpredicts the true curve, as might be expected from the $g_*$ estimation difficulties identified previously. Once more, if higher thresholds are used for fitting the new model then the bias of $\eta$ will be reduced and the extrapolation properties of the new model improved accordingly.
Figure 5.4: Joint density estimates obtained using the new logistic model for a) data set A; b) data set B; c) data set C and d) data set D. The contours are at $10^{-j/2}$ for $j = 4, ..., 11$. 
Figure 5.5: Extrapolated contours of the survivor function $\bar{F}(x, y) = p$ for the true model (line) and for the fitted logistic survivor model $\bar{F}_{XY}$ (dashed) with data points superimposed for a) data set A; b) data set B; c) data set C and d) data set D. The contour levels are at $p = 10^{-i}$ for $i = 2, \ldots, 7$. 
Chapter 5. Asymptotically independent joint tail modelling in practice

Figure 5.6: Scatter plots of the two environmental data sets with the selected thresholds included. The data sets are: a) data set I and b) data set II.

5.5 Application to real data

In order to analyse and illustrate the new method further we consider here two sets of bivariate environmental data.

Data set I: wave-surge data of Coles and Tawn (1994), consisting of 2894 approximately independent events which occurred during 1971-77 in Cornwall.

Data set II: rain-wind data of Anderson and Nadarajah (1993), consisting of 1737 independent pairs of rainfall and wind speed values measured at Eskdalemuir during 1970-86.

Scatter plots of both data sets are given in Figure 5.6.

Estimating the joint tail of these data requires not only estimating the extremal dependence structure but also estimation of the marginal distributions, since these are unknown. In this section, we demonstrate how these features can be estimated simultaneously using the maximum likelihood methodology described in Section 5.2.
Chapter 5. Asymptotically independent joint tail modelling in practice

According to Ledford and Tawn (1997), appropriate thresholds may be obtained by transforming the marginal variables, \((X^*, Y^*)\) say, to unit Fréchet variables \((X, Y)\), by using empirical probability integral transformations, and choosing the threshold \(u\) at the empirical 90% point of the structure variable \(V = \min(X, Y)\). The resulting threshold \(u\) has values 3.70 and 3.28 for the wave-surge and rain-wind transformed data, with 290 and 173 joint threshold exceedances, respectively. The corresponding thresholds \(u_1\) and \(u_2\) for the original variables are the image of the threshold \(u\) under the marginal transformations and take values \(u_1 = 3.76\) and \(u_2 = 0.15\) for the wave-surge data, and \(u_1 = 10.1\) and \(u_2 = 21\) for the rain-wind data. These values are also represented in Figure 5.6.

5.5.1 Inference and results

Firstly, an initial analysis of just the dependence features of the two data sets is carried out using the unit Fréchet marginally distributed transformed data as defined above. We start by investigating whether our asymptotically independent joint tail model is preferable here to models based on asymptotic dependence within the joint tail using the transformed data and result (5.11) to test asymptotic independence. For the likelihood function described in equations (5.8), the likelihood ratio statistic has values 2.7 and 53.7 for the wave-surge and rain-wind transformed data, respectively. The value for the rain-wind case is clearly significant, indicating that the marginal extremes of rain and wind exhibit asymptotic independence. For the wave-surge case though, the statistical value is not significant which suggests that asymptotic dependence between the extremes should not be rejected. Diagnostic-based estimates of \(\eta\) obtained using the chosen threshold \(u\) and the structure variable \(V\), as defined above, with associated standard errors (in parentheses) are 0.85 (0.11) and 0.63 (0.12) for data sets I and II, respectively. These results confirm the conclusions above and also suggest that extremes of the marginal variables are positively associated for both data sets.

In order to see if a symmetric joint tail model is appropriate for our data, or if an asymmetric one is needed, we use result (5.15) and test symmetry between extremes. The likelihood ratio statistic has values 14.6 and 8.28 for the wave-surge and rain-wind transformed data, respectively, suggesting there is non-symmetry for both data sets. Similar conclusions can be drawn by using the original data and the above tests defined for the full parameter likelihood functions described in Section 5.2. Testing ray
independence is not appropriate here since symmetry is rejected for both data sets and, as pointed out in Section 5.3.3, ray independence implies symmetry.

Since the latter test suggested the presence of some asymmetry in both data sets we use and compare both the modified logistic and asymmetric logistic joint tail models throughout this analysis. The first stage of the model fitting is to estimate the dependence and marginal parameters of our model for the chosen thresholds. Initially, we estimate separately the dependence and marginal parameters and then undertake joint estimation of those parameters simultaneously.

Dependence parameter estimates obtained by fitting both the modified logistic and asymmetric logistic models to unit Fréchet transformed data and to the original data are given in Table 5.2 using the chosen threshold $u_i$ together with associated standard errors (in parentheses). Obtaining these values required maximising the four likelihood functions mentioned in Section 5.2. Estimates of the generalised Pareto distribution scale and shape parameters are given in Table 5.3 for each variable for separate and joint analyses. For the separate analyses we used the thresholds $u_1$ and $u_2$ defined above for each data variable. The starting values used in the simultaneous estimation of marginal and dependence parameters were the values provided by the separate analyses and those obtained by fitting the model to the unit Fréchet transformed data.

<table>
<thead>
<tr>
<th>Data</th>
<th>$\eta$</th>
<th>$\alpha$</th>
<th>$\hat{\phi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_1$</td>
<td>$F_2$</td>
<td>$J_1$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\eta}_1$</td>
<td>$\hat{\eta}_2$</td>
<td>$\hat{\alpha}_1$</td>
</tr>
<tr>
<td>I</td>
<td>0.94</td>
<td>0.95</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>II</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.2: Dependence parameter estimates: Estimates of the dependence parameters for each data set obtained by fitting the model to the unit Fréchet empirically transformed data, using the modified logistic ($F_1$) and asymmetric logistic ($F_2$) models, and by fitting the model to the original data, using the full logistic ($J_1$) and asymmetric logistic ($J_2$) models. Standard errors obtained by the delta-method are given in parentheses. The thresholds used were the 90% point of the empirical structure variable distribution in each case.
Table 5.3: Marginal parameter estimates: Estimates of the GPD shape and scale parameters for each marginal data set obtained through separate univariate analyses (S), and joint analyses using the full modified logistic (J₁) and asymmetric logistic (J₂) models. Standard errors obtained by the delta-method are given in parentheses.

Table 5.2 shows that dependence parameter estimates obtained by fitting our models to the empirically transformed data or to the original data, whose margins are then estimated by a GPD, are similar and have similar standard errors. It is also clear that the estimates of the coefficient of tail dependence given in this table are larger than those obtained previously from the η-diagnostic and all have smaller standard errors. In the same table, estimates of the asymmetry parameter q suggest the existence of asymmetry for both data sets. Table 5.3 also show similar results for the marginal parameter estimates and their associated standard errors obtained through separate univariate and the joint analyses. The similarity of the marginal parameter estimates for the separate and joint analyses suggests that the fitted dependence structures are representative of the observed extremal dependence.

As mentioned before, the adequacy of the fitted model as well as the chosen threshold can be checked by plotting the model estimates of each parameter for a range of thresholds. Figure 5.7 shows dependence parameter estimates calculated using the modified asymmetric logistic model for a range of V-variable threshold probabilities for data sets I and II. The figure also includes 95% confidence intervals based on the standard errors. Note that since η ≤ 1, the coverage of the confidence interval for η is not exactly 95% when the estimates of η are close to 1.

Relative stability in the parameter estimates can be observed for thresholds above the chosen structure variable threshold probability 0.9. For data set I, the η = 1 line, corresponding to asymptotic dependence, is close to or contained within the confidence interval for η over the range of thresholds used. In contrast, for data set II, the
Figure 5.7: Estimates of the dependence parameters $\eta$ (top), $\alpha$ and $\varrho$ (bottom) of the modified asymmetric logistic model for a range of structure variable threshold probabilities for data set I (left) and data set II (right), together with 95% confidence intervals based on the standard errors. The lines $\eta = 0.5$, $\eta = 1$ and $\varrho = 1$ are included for reference.
Chapter 5. Asymptotically independent joint tail modelling in practice

Confidence interval is more consistent with asymptotic independence. Figure 5.7 also indicates that estimates of \( \alpha \) are smaller than twice the value of the model estimate of \( \eta \), suggesting concave ray dependence for both data sets. This would be supported by a visual check of plots of the quantity \( \alpha - 2\eta \) (not reported here). Plots of the asymmetry parameter estimates show that the \( q = 1 \) line, corresponding to symmetry, is not contained in the confidence interval over nearly the whole range of thresholds used for data set II. For data set I, although the line is contained in the confidence interval for high thresholds, there is some evidence of asymmetry. Corresponding plots of the estimates of the dependence parameters \( \alpha \) and \( \eta \) as the threshold varies, calculated using the modified logistic model are very similar to those in Figure 5.7 and therefore not reported here.

Plots of the estimated ray dependence function \( g_* \) obtained from the diagnostic in equation (5.16) using the diagnostic estimates of \( \eta \) given previously are shown in Figure 5.8 for data sets I and II, together with approximate 95% delta-method based pointwise confidence interval. The fitted ray dependence functions obtained using the full modified logistic and asymmetric logistic models are also given. Strong ray dependence is apparent for data set I, whereas the less pronounced curvature for data set II suggests a weaker ray dependence. Another clear feature of the ray dependence for both data sets is asymmetry and the fitted ray dependence function for the asymmetric model has substantially better agreement with the diagnostic estimated \( g_* \) than the estimate obtained using the symmetric model. This confirms that the use of an asymmetric joint tail model is appropriate for both data sets.

Since a fully parametric model for the joint distribution function is available for the region where both of the original marginal variables exceed their respective thresholds, appropriate differentiation yields the joint density function of the fitted model. Contours of the joint density functions for the full modified asymmetric logistic model are plotted in Figure 5.9 for each data set. These density contour plots show strong extremal dependence in data set I and weak dependence in data set II.
Chapter 5. Asymptotically independent joint tail modelling in practice

Figure 5.8: Diagnostic estimates (-----) of \( g^w(w) \) with approximate pointwise 95% delta-method based confidence intervals (\( \cdots \)) and fitted ray dependence function estimates obtained from joint analyses using the full logistic (---) and asymmetric logistic (---) models for the a) wave-surge and b) rain-wind data sets.

5.5.2 Extrapolation

Finally, we investigate in this section how the fitted model extrapolates from the sample information. Joint tail contour curves from the full asymmetric logistic survivor model \( \bar{F}_{X \cdot Y} \) as in equation (5.7) are plotted in Figure 5.10 for data sets I and II. The contour levels are at \( 10^{-i} \) for \( i = 2, \ldots, 7 \), that is, we extrapolate up to \( p = \bar{F}_{X \cdot Y}(x^*, y^*) = 10^{-7} \), where \((X^*, Y^*)\) represents the original data.

There is good agreement between the data points and the fitted curves for both cases confirming once again the good performance of the fitted asymmetric logistic model. Again, the survivor function contour plots reflect the strong dependence in data set I and the weaker dependence in data set II. Similar results are obtained if the logistic survivor model defined in equation (5.6) is used instead of the asymmetric logistic one.
Chapter 5. Asymptotically independent joint tail modelling in practice

Figure 5.9: Joint density estimates obtained using the full asymmetric logistic model for a) data set I and b) data set II. The contours are at $10^{-j/2}$ for $j = 2, \ldots, 8$ and $j = 6, \ldots, 13$ for data sets I and II, respectively.

Figure 5.10: Extrapolated joint tail contour curves of the full asymmetric logistic survivor model with data points superimposed for a) data set I and b) data set II. The contours are at $10^{-j}$ for $j = 2, \ldots, 7$. 

115
5.5.3 Comments

Our analysis of the two environmental data sets demonstrates that both ray dependence and tail dependence are much stronger in the wave-surge data than in the rain-wind data, and also that models based on asymptotic dependence (i.e. $\eta = 1$) may be inappropriate for both data sets. Indeed, the use of our models has advantages over existing methods even for data sets where there is strong extremal dependence, like the wave-surge data, since they provide more flexibility for joint tail modelling than classical BEV methods which are only really appropriate for componentwise maxima data. Although both the modified logistic and asymmetric logistic models performed well for the two data sets, the performance of the asymmetric model is considerably better.

5.6 Conclusion

This chapter shows that our new joint tail models provide substantial improvements over existing models. The focus is on modelling joint tails rather than distributions of componentwise maxima. Consequently, our models are more relevant for applications where usually data consist of joint observations rather than being of componentwise maxima form. Our new models perform well for simulated and real data and provide a unified framework covering both asymptotic independence and asymptotic dependence, being satisfactory even for cases of weak or negative association. We have developed a likelihood based framework for simultaneous estimation of joint and marginal parameters, and have developed likelihood based tests that are useful within model selection.
Chapter 6

Point process results for asymptotic independence

Statistical techniques for analysing multivariate extremes are often based on the de Haan (1985) point process representation described in Section 1.4.2, see for example Coles and Tawn (1991, 1994). As was seen in Section 1.4.3, the limiting distribution of the normalised componentwise maxima of iid unit Fréchet variables may be derived from this point process and has the form $\exp(-V)$ where $V$ is the dependence function defined in equation (1.14). As discussed previously, the dependence structures accommodated by this approach all have $\eta = 1$, corresponding to asymptotic dependence, or have $\eta = 1/2$ when $V$ corresponds to exact independence. In no cases does an asymptotically independent limit result hold as all asymptotically independent cases are degenerate and lead to exact independence.

The de Haan (1985) point process was extended by Ledford and Tawn (1997) who developed a point process representation for asymptotic independence. However, their examination of the new point process was quite limited, no conditions detailing the required properties of the intensity were derived and no parametric examples were provided. In this chapter we study their point process, develop an analogue of the standard componentwise maxima result for the asymptotically independent case, and provide some parametric examples.
Chapter 6. Point process results for asymptotic independence

6.1 Introduction

Let \((X_i, Y_i), \ldots, (X_n, Y_n)\) denote iid bivariate random variables with unit Fréchet margins and joint survivor function satisfying equation (4.1), and let \(b_n\) satisfy \(nP(b_n, b_n) = 1\). Then, \(b_n = o(n)\), and commonly \(b_n = O(n^\alpha)\), unless the componentwise maxima are asymptotically dependent, in which case \(b_n = O(n)\).

Our main focus is the sequence of point processes given by Ledford and Tawn (1997) defined by

\[ P^*_n = \left\{ \left( \frac{X_i}{b_n}, \frac{Y_i}{b_n} \right) : i = 1, \ldots, n \right\}. \tag{6.1} \]

We remark that the normalising constants \(b_n\) in \(P^*_n\) are lighter than those for the point process \(P_n\) defined in Section 1.4.2 since they are appropriate to stabilise joint tail observations rather than the componentwise maxima. Ledford and Tawn (1997) show that \(P^*_n \Rightarrow P^*\) as \(n \to \infty\) where \(P^*\) is a non-homogeneous Poisson process on \(\mathbb{R}_+^2 \setminus \{0 \cup (\mathbb{R}_+ \times 0) \cup (0 \times \mathbb{R}_+)\}\) with point intensity given by

\[ \mu_\eta(dr \times dw) = r^{-(1+1/\eta)}h_\eta(w) \, dr \, dw, \tag{6.2} \]

where \(R = X + Y\) and \(W = X/Y\) are the pseudo-polar coordinates and the function \(h_\eta\) is a non-negative measure density\(^1\) on \([0,1]\). Our first result is to show that the angular measure density given in equation (6.2) coincides precisely with that given in equation (4.6). Thus, from our previous results, we are able to derive both theoretical conditions for obtaining suitable models for the point-process angular measure density and a range of parametric examples.

To verify that the measure density defined in (6.2), say \(h^*_\eta\), and the measure density \(h_\eta\) defined in (4.6) coincide, we use a very similar argument to that which gives the Generalised Pareto distribution from the univariate point process result (see Section 1.3.5). Denoting the intensity measure of the point process (6.1) by \(\Lambda_\eta\), we have that the joint survivor function of the bivariate random variable \((S, T)\) as defined in equation (4.3) is given by

\[ \Pr(S > x, T > y) = \lim_{n \to \infty} \frac{\Pr(X > b_n x, Y > b_n y)}{\Pr(X > b_n, Y > b_n)} = \Lambda_\eta \{(x, \infty) \times (y, \infty)\}, \tag{6.3} \]

\(^1\)Function \(h_\eta\) is such that \(h_\eta(w) = \frac{dH_\eta(w)}{dw}\) if the measure \(H_\eta\) is differentiable; atomic masses are considered otherwise.
since \( \Lambda_{\eta} \{ (1, \infty) \times (1, \infty) \} = 1 \) as we have \( nF(b_n, b_n) = 1 \). Taking the mixed second derivative of both end terms in (6.3) with respect to \( x \) and \( y \), using equations (4.6) and (6.2) and transforming to \((R, W)\) coordinates, we obtain

\[
(x + y)^{-(1+1/n)} h_{\eta} \{ x/(x + y) \} = (x + y)^{-(1+1/n)} h_{\eta}^* \{ x/(x + y) \},
\]

which establishes that \( h_{\eta}(w) = h_{\eta}^*(w) \) for every \( w \in (0,1) \).

### 6.2 Distribution of the componentwise maxima \((M_{X,n}, M_{Y,n})\) out of those points which are simultaneously large

In this section we study the bivariate analogue of result (1.14), concerning the limiting distribution of the componentwise maxima, for the case of asymptotic independence. This topic was examined briefly by Ledford and Tawn (1997), although careful inspection of their treatment suggests it is incomplete and a more detailed approach is required. This is addressed here. The derivation presented in this section is based on the implicit assumption that considering a sequence of sets \( A_n \) converging to the set \( A \) and a sequence of point processes such that \( \mathcal{P}_n \xrightarrow{w} \mathcal{P} \), where \( \mathcal{P} \) is a point process, then \( \mathcal{P}_n(A_n) \xrightarrow{w} \mathcal{P}(A) \). We consider then the point process \( \mathcal{P}_n^* \), defined in equation (6.1), but restrict attention to the domain \( R_\theta \) where

\[
R_\theta = \{(x, y) \in \mathbb{R}_+^2 : x > \theta, y > \theta\}
\]

for some fixed \( 0 < \theta < 1 \). By doing this, we look at a domain bounded away from the axes. For any fixed \( \theta > 0 \), it is straightforward to show that \( \mathcal{P}_n^* \xrightarrow{w} \mathcal{P}_\theta^* \) on \( R_\theta \) by Kallenberg's Theorem (Resnick, 1987, Proposition 3.22), where \( \mathcal{P}_\theta^* \) is a non-homogeneous Poisson process defined on \( R_\theta \) with point intensity (6.2). Our objective is to obtain the limiting distribution of the normalised componentwise maxima \((M_{X,n}, M_{Y,n})\) for points simultaneously large.

First we consider the region \( R_{\theta b_n} = \{(x, y) \in \mathbb{R}_+^2 : x > \theta b_n, y > \theta b_n\} \). Then, for any fixed \( \theta > 0 \) and for \( x, y > 0 \),

\[
\lim_{n \to \infty} \Pr \left( M_{X,n} < b_n x, M_{Y,n} < b_n y \middle| \text{of points in } R_{\theta b_n} \right) = \exp \{-V_\theta(x, y)\} \quad (6.4)
\]
where $V_0(x, y)$ is the expected number of points of $\mathcal{P}_0^*$ in $A = R_0 \setminus \{[\theta, x] \times [\theta, y]\}$.

Exploiting the point intensity (6.2), we may write

$$V_0(x, y) = \Lambda_0 \{(x, \infty), (\theta, \infty)\} + \Lambda_0 \{(\theta, \infty), (y, \infty)\} - \Lambda_0 \{(x, \infty), (y, \infty)\}$$

$$= \eta \int_0^1 \left\{ \min \left( \frac{w}{x}, \frac{1}{\theta} \right) \right\}^{1/\eta} h_\eta(w) \, dw + \eta \int_0^1 \left\{ \min \left( \frac{w}{\theta}, \frac{1}{y} \right) \right\}^{1/\eta} h_\eta(w) \, dw$$

$$- \eta \int_0^1 \left\{ \min \left( \frac{w}{x}, \frac{1}{y} \right) \right\}^{1/\eta} h_\eta(w) \, dw, \quad (6.5)$$

where the integration is over the open interval $0 < w < 1$.

Clearly, as $\theta \to 0^+$ we obtain $\mathcal{P}_0^* \Rightarrow \mathcal{P}^*$. Taking the limit as $\theta \to 0^+$ in equation (6.5) will provide a definition of the expected number of points of $\mathcal{P}^*$ in region $\mathbb{R}_+^2 \setminus \{(0, x) \times (0, y)\}$ in terms of the measure density $h_\eta$. However, attention has to be paid first to the case where one or both of those integrals containing $\theta$ in equation (6.5) have infinite limits as $\theta \to 0^+$. If this is the case then as $\theta \to 0^+$ the limiting probability in (6.4) is degenerate (i.e. zero) and the componentwise maxima of points in the domain of $\mathcal{P}^*$ are not stabilised by the normalising constant $h_\eta$. Therefore, in order to obtain a non-degenerate limit, a heavier normalising constant is required. However, the choice of that constant is not addressed here since our purpose is to clarify the Ledford and Tawn (1997) result. In any case, the end result would not be very interesting as even if a non-degenerate limit is obtained by suitable choice of norming, the resulting limit distribution has exactly independent margins. Additionally, the case where the integrals do not converge as $\theta \to 0^+$ corresponds to the marginal variables (of the points in the domain of $\mathcal{P}^*$) having a heavier tail than the joint tail, which is known to have index $\eta$. Thus our approach is to focus on the behaviour of the given point process, and restrict attention to the non-degenerate case that arises when these integrals do converge. Taking the limit as $\theta \to 0^+$ in equation (6.5) we conclude that

$$\lim_{\theta \to 0^+} V_0(x, y) = \eta \int_0^1 \left( \frac{w}{x} \right)^{1/\eta} h_\eta(w) \, dw + \eta \int_0^1 \left( \frac{1-w}{y} \right)^{1/\eta} h_\eta(w) \, dw$$

$$- \eta \int_0^1 \left\{ \min \left( \frac{w}{x}, \frac{1-w}{y} \right) \right\}^{1/\eta} h_\eta(w) \, dw$$

$$= \eta \int_0^1 \left\{ \max \left( \frac{w}{x}, \frac{1-w}{y} \right) \right\}^{1/\eta} h_\eta(w) \, dw.$$

Thus, defining

$$V_\eta(x, y) = \lim_{\theta \to 0^+} V_0(x, y) = \eta \int_0^1 \left\{ \max \left( \frac{w}{x}, \frac{1-w}{y} \right) \right\}^{1/\eta} h_\eta(w) \, dw, \quad (6.6)$$
where the integration is over the open interval \(0 < w < 1\), we obtain

\[
\lim_{s \to 0^+} \lim_{n \to \infty} \Pr \left( M_{X,n} < b_n x, M_{Y,n} < b_n y \right) \text{ of points in } R_{\theta_n} = \exp \{-V_\eta(x,y)\} = G_\eta(x,y). \quad (6.7)
\]

Note that the limiting results given above are degenerate but still have valid interpretation for the infinite integrals case, since they correctly give the limiting probability zero. This latter result gives the joint distribution of componentwise maxima for pairs of variables which are simultaneously large and extends the standard componentwise maxima result to the asymptotically independent case. It also re-states the Ledford and Tawn (1997) result concerning only points in the region \(R_n = \{(x, y) \in \mathbb{R}^2 : x > n, y > n\}\) and clarifies the convergence issues. Clearly, the dependence function \(V_\eta\) defined in (6.6) is homogeneous of order \(-1/\eta\), i.e. \(V_\eta(tx, ty) = t^{-1/\eta}V_\eta(x, y)\) for all \(t > 0\). Thus \(G^\eta_n(n^nx, n^ny) = G_\eta(x, y)\) and \(G_\eta\) is max-stable.

Result (6.7) concerns points within the limiting point process \(P^\ast\). If this restriction is omitted, then we are back in the classical BEV case and the existing results hold. Thus the results given in Section 3.2.1 of Kotz and Nadarajah (2000) appear to be either incomplete or incorrectly stated.

The current derivation of equation (6.7) based on joint survivor type regions leads to a result valid for \(w\) in the open interval \((0, 1)\), and therefore, a result where masses at the boundaries \(w = 0\) and \(w = 1\) play no role. However, it might be useful to extend the definition of the appropriate integral to the closed interval \([0, 1]\) as the results that then follow have a more direct relationship to the existing results for bivariate extremes. This extension of the domain to include the boundaries is only possible for cases where the required integrals converge and provided that only finite masses are added at the end points.

### 6.2.1 Obtaining \(h_\eta\) from \(V_\eta\)

In equation (6.6) the dependence function \(V_\eta\) is defined in terms of a given measure density \(h_\eta\). We investigate here how the density \(h_\eta\) can be obtained from a given \(V_\eta\). The approach we use is similar to that in Coles and Tawn (1991) and, indeed, their result is a special case of that presented here.

Using equation (4.7) and the definition of \(V_\eta\), the survivor function \(F_{ST}\) can be writ-
ten as
\[ F_{ST}(x, y) = V_\eta(x, \infty) + V_\eta(\infty, y) - V_\eta(x, y). \]  
(6.8)

Then, taking the mixed derivative in equation (6.8) with respect to \( x \) and \( y \), we obtain
\[ \frac{\partial^2 V_\eta}{\partial x \partial y} = -r^{-(2+1/n)} h_\eta(w) \]
where \( r = x + y \) and \( w = x/(x + y) \). This provides an extension of result (4.28) of Coles and Tawn (1991) to the asymptotically independent case, provided the density \( h_\eta \) exists. However, the Coles and Tawn (1991) result is more extensive since a similar relationship is valid also for masses on the boundaries.

### 6.3 Distribution of the componentwise maxima \((M_{S,n}, M_{T,n})\)

Consider the iid random variables \((S_1, T_1), \ldots, (S_n, T_n)\) with common distribution function \(F_{ST}\), as previously defined in equation (4.3). We derive here the limit distribution of the componentwise maxima \((M_{S,n}, M_{T,n})\). As may be expected, the results here are those suggested by classical BEV theory.

Using the obvious notation
\[
\lim n \Pr (M_{S,n} < n^\eta x, M_{T,n} < n^\eta y) = \lim n \Pr (S > n^\eta x, T > n^\eta y) - \Pr (S > n^\eta x, T > n^\eta y) = \exp \{ -V_\eta(x, y) \} \equiv G_\eta(x, y), 
\]
(6.9)
for \( x, y > 0 \), since, by equation (4.7),
\[
\lim n \{ \Pr (S > n^\eta x) + \Pr (T > n^\eta y) - F_{ST}(n^\eta x, n^\eta y) \} =
\]
\[
= \lim n \eta \left[ \int_0^1 \{ \min \left( \frac{w}{n^\eta x}, 1 - w \right) \}^{1/\eta} h_\eta(w) \, dw \right]
= \eta \int_0^1 \{ \min \left( \frac{w}{x}, 1 - w \right) \}^{1/\eta} h_\eta(w) \, dw.
\]
(6.10)
That is, the normalised componentwise maxima of the points \((S_i, T_i)_{i=1,\ldots,n}\) has limiting distribution \(G_\eta = \exp(-V_\eta)\) where \(V_\eta\) is the dependence function defined in equation (6.6). Note that, again, the only case considered is where the integrals containing \(n^\eta\) in equation (6.10) converge. As in the previous section, the case where the
Chapter 6. Point process results for asymptotic independence

integrals do not both converge corresponds to tails of the margins of \((S,T)\) being heavier than the joint tail and to the value of the limiting probability in equation (6.9) being zero.

6.4 Examples of \(V_\eta\)

We give here some parametric examples of the dependence function \(V_\eta\). These are calculated using equation (6.6) for the extended closed interval \([0,1]\) and the measure densities of the modified logistic, asymmetric logistic and mixed models defined in Examples A, B and C, respectively, of Section 4.3.

A. Consider the modified logistic model, as in Example A, where

\[
h_\eta(w) = \frac{\eta - \alpha}{\alpha \eta^2 (2 - 2^{\alpha/\eta})} \left\{ w^{-1/\alpha} + (1 - w)^{-1/\alpha} \right\}^{-2 + \alpha/\eta} \{w(1 - w)\}^{-(1+1/\alpha)}.
\]

Then, elementary integration yields

\[
V_\eta(x, y) = \begin{cases} 
(2 - 2^{\alpha/\eta})^{-1} (x^{-1/\alpha} + y^{-1/\alpha})^{\alpha/\eta} & \text{for } \alpha < \eta, \\
+\infty & \text{for } \alpha \geq \eta.
\end{cases}
\]

Thus, by equation (6.7), we obtain

\[
\lim_{\theta \to \theta^+} \lim_{n \to \infty} \Pr \left( M_{X,n} < b_n x, M_{Y,n} < b_n y \right) \quad \text{of point in } R_{\theta b_n} = \begin{cases} 
\exp \left\{ -(2 - 2^{\alpha/\eta})^{-1} (x^{-1/\alpha} + y^{-1/\alpha})^{\alpha/\eta} \right\} & \text{for } \alpha < \eta, \\
0 & \text{for } \alpha \geq \eta.
\end{cases}
\]

Similarly, the normalised componentwise maxima \((M_{S,n}/n^\eta, M_{T,n}/n^\eta)\) can be shown to have this same limiting distribution. This implies that if \(\alpha \geq \eta\) the probability of all points \((S_i, T_i)\) being in the region \([1, n^\eta s) \times [1, n^\eta t)\) converges to zero when \(n \to \infty\). From a geometric viewpoint, this means that componentwise maxima tend to occur close to the boundary of the domain \((s = 1\) and \(t = 1)\) and the resulting componentwise maxima are independent in the limit. In fact, when \(\alpha\) increases points become closer to these boundaries as can be seen from Figure 7.1 in Chapter 7. To further clarify this result, we note, as seen in Section 4.4, that the marginal distributions of the variable \((S, T)\) have tails with shape parameter \(\alpha\) when \(\alpha \geq \eta\), and are therefore heavier than the joint tail of \((S, T)\) which has shape parameter \(\eta\).
B. The modified asymmetric logistic model in Example B has measure density

\[ h_\eta(w) = \frac{n - \alpha}{\alpha n^2 \theta \phi N_{\theta \phi}} \left\{ \left( \frac{w}{\theta} \right)^{-1/\alpha} + \left( \frac{1 - w}{\phi} \right)^{-1/\alpha} \right\}^{\alpha/n - 2} \left\{ \frac{w}{\theta} \left( 1 - \frac{w}{\phi} \right) \right\}^{-(1+1/\alpha)} \]

for 0 < w < 1 and atoms of mass \((1 - \phi^{1/\eta})/(\eta N_{\theta \phi})\) and \((1 - \theta^{1/\eta})/(\eta N_{\theta \phi})\) at \(w = 0\) and \(w = 1\), respectively. Proceeding as above, we obtain

\[ V_\eta(x, y) = \begin{cases} N_{\theta \phi}^{-1} \left\{ \left( \frac{\phi}{\theta} \right)^{-1/\alpha} + \left( \frac{\theta}{\phi} \right)^{-1/\alpha} \right\}^{\alpha/\eta} + \frac{1 - \theta^{1/\eta}}{x^{1/\eta}} + \frac{1 - \phi^{1/\eta}}{y^{1/\eta}} \quad &\text{for } \alpha < \eta, \\
+\infty &\text{for } \alpha \geq \eta. \end{cases} \]

The limiting distribution of the normalised componentwise maxima in equations (6.7) and (6.9) follow similarly to the previous case and the same conclusions can be made for the \(\alpha \geq \eta\) case.

C. The modified mixed model in Example C has measure density

\[ h_\eta(w) = 2^{1/\eta} \eta^{-2} (\eta + 1), \]

for 0 < w < 1 and atoms of equal mass \(2^{1/\eta}(1 - \theta)/(\eta \theta)\) at \(w = 0\) and \(w = 1\).

Integrating as before we obtain

\[ V_\eta(x, y) = 2^{1/\eta} \theta^{-1} \left\{ x^{-1/\eta} + y^{-1/\eta} - \frac{\theta}{(x + y)^{1/\eta}} \right\}. \]

### 6.5 Marginal properties

In this section marginal properties associated with the joint distribution defined in equation (6.7) will be studied.

The limiting result in (6.7) gives the joint distribution of the normalised componentwise maxima of points in the domain of \(P^*\). From this result, the marginal behaviour is given by

\[
\lim_{\theta \to 0^+ \eta \to \infty} \lim_{n \to \infty} \Pr(M_{X,n} < b_n x) = \exp \left\{ -V_\eta(1, \infty) x^{-1/\eta} \right\}
\]

and

\[
\lim_{\theta \to 0^+ \eta \to \infty} \lim_{n \to \infty} \Pr(M_{Y,n} < b_n y) = \exp \left\{ -V_\eta(\infty, 1) y^{-1/\eta} \right\}. \quad (6.11)
\]

Thus, the margins have a Fréchet distribution with shape and scale parameters equal to \(\eta\) and \(V_\eta(1, \infty) = \eta \int_0^1 w^{1/\eta} h_\eta(w) \, dw\) or \(V_\eta(\infty, 1) = \eta \int_0^1 (1 - w)^{1/\eta} h_\eta(w) \, dw\), respectively. So, unlike in the standard BEV case, the margins here are not unit Fréchet distributed.
Chapter 6. Point process results for asymptotic independence

Examples:

A. For the modified logistic model in Example A, the scale parameter \( V_\eta(1, \infty) \) is equal to \( (2 - 2^{\alpha/\eta})^{-1} \) if \( \alpha \leq \eta \) whereas the result is degenerate if \( \alpha > \eta \). Thus

\[
\lim_{d_\theta \to 0^+} \lim_{n \to \infty} \Pr \left( M_{X,n} < b_n x \right) = \begin{cases} 
\exp \left\{ - (2 - 2^{\alpha/\eta})^{-1} x^{-1/\eta} \right\} & \text{for } \alpha \leq \eta, \\
0 & \text{for } \alpha > \eta.
\end{cases}
\]

By symmetry, the same result holds for the other margin.

B. For the modified asymmetric logistic model in Example B, both scale parameters for each margin in equations (6.11), \( V_\eta(1, \infty) \) and \( V_\eta(\infty, 1) \), are equal to \( N_{\eta \phi}^{-1} \) if \( \alpha \leq \eta \) whereas if \( \alpha > \eta \) the result is degenerate. Thus both margins satisfy

\[
\lim_{d_\theta \to 0^+} \lim_{n \to \infty} \Pr \left( M_{X,n} < b_n x \right) = \begin{cases} 
\exp \left\{ -N_{\eta \phi}^{-1} x^{-1/\eta} \right\} & \text{for } \alpha \leq \eta, \\
0 & \text{for } \alpha > \eta.
\end{cases}
\]

C. The modified mixed model of Example C yields \( V_\eta(1, \infty) = 2^{1/\eta} \theta^{-1} \) and thus has marginal behaviour

\[
\lim_{d_\theta \to 0^+} \lim_{n \to \infty} \Pr \left( M_{X,n} < b_n x \right) = \exp \left\{ -2^{1/\eta} \theta^{-1} x^{-1/\eta} \right\}.
\]

Again, since this model is symmetric, the same holds for the other margin.

6.6 Comments

As studied above, the limiting distribution \( G_\eta \) has Fréchet margins with shape parameter \( \eta \) and scale parameters \( V_\eta(1, \infty) \) or \( V_\eta(\infty, 1) \). By adopting different normalising conditions it is of course possible to obtain a limiting distribution \( G_\eta^* \) that has Fréchet margins with the same shape parameter but scale parameters equal to 1. Specifically the point process

\[
P_n^# = \left\{ \left( \frac{X_i}{V_\eta^*(1, \infty) b_n}, \frac{Y_i}{V_\eta^*(1, \infty) b_n} \right) \: i = 1, \ldots, n \right\}
\]

where \( 0 < V_\eta(1, \infty) < \infty \), achieves this for parametric models such that \( V_\eta(1, \infty) = V_\eta(\infty, 1) \), in particular, for all our parametric examples and all symmetric parametric models. It follows that \( P_n^# \overset{d}{\to} P_n^\# \) as \( n \to \infty \) where \( P_n^\# \) is a non-homogeneous Poisson process on \( \mathbb{R}^2 \setminus \{ \emptyset \cup (\mathbb{R}^+ \times 0) \cup (0 \times \mathbb{R}^+) \} \) with point intensity

\[
\mu_\eta^\# (dr \times dw) = r^{-(1+1/\eta)} h_\eta^\# (w) \: dr \: dw.
\]

125
Chapter 6. Point process results for asymptotic independence

It is straightforward to see that \( h_\eta^*(w) = V^{-1}_\eta(1, \infty)h_\eta(w) \) with \( h_\eta \) as in equation (6.2), and thus the dependence function

\[
V_\eta^*(x, y) = \eta \int_0^1 \left\{ \max \left( \frac{w}{x}, \frac{1 - w}{y} \right) \right\}^{1/\eta} h_\eta^*(w) \, dw = V^{-1}_\eta(1, \infty)V_\eta(x, y)
\]

is such that \( V_\eta^*(1, \infty) = 1 \). We conclude that the corresponding distribution function \( G_\eta^*(x, y) = \exp\left\{ -V_\eta^*(x, y) \right\} \) has Fréchet margins with shape parameter \( \eta \) and scale parameter \( V_\eta^*(1, \infty) = 1 \).

In this chapter point process results were derived only for the bivariate case. These results may be extended to the multivariate case to obtain a corresponding d-dimensional case. Suppose that \( X_1, \ldots, X_n \) are iid d-dimensional random variables with joint distribution \( F(x) = F(x_1, \ldots, x_d) \), denote the d-vector of the componentwise maxima by \( M_n \) and define \( R_{\theta b_n} = \{ x \in \mathbb{R}^d : x_1 > \theta b_n, \ldots, x_d > \theta b_n \} \). Then, using the obvious notation, we have

\[
\lim_{\theta \to 0^+} \lim_{n \to \infty} \Pr \left( M_n < b_n x \right. \left. \text{ of points in } R_{\theta b_n} \right) = \exp \left[ -\eta \int_{S_{d-1}} \left\{ \max_{1 \leq i \leq d} \left( \frac{w_i}{x_i} \right) \right\}^{1/\eta} h_\eta(w) \, dw \right]
\]

where \( S_d \) is the \((d-1)\)-dimensional unit simplex. The extension of result (6.9) to the d-dimensional case follows similarly.
Chapter 7

Simulation methods

In this chapter, methods for simulating points from the symmetric and asymmetric bivariate logistic distribution function $F_{ST}$, defined in Section 1.4.4, as well as methods to generate points from the symmetric and asymmetric bivariate logistic distribution function $G_{\eta}$, as in Section 6.4, are given. The approach considered here uses transformations to derive random variables with a joint distribution from which simulation is straightforward. Methods for simulating from these distributions are vital for undertaking Monte Carlo integration when calculating expectations with respect to the underlying model, and also are useful for testing modelling and estimation approaches on simulated data.

7.1 Simulation from the distribution $F_{ST}$

In this section, methods for simulating from the distribution function $F_{ST}$ are developed for the modified logistic and asymmetric logistic joint tail models.

7.1.1 Using the modified logistic model

The density function of the modified logistic model is given by

$$f_{ST}(s,t) = \frac{\eta - \alpha}{\alpha \eta^2 (2 - 2^\alpha/\eta)} \left(s^{-1/\alpha} + t^{-1/\alpha}\right)^{\alpha/2 - 2}(st)^{-(1+1/\alpha)} \quad \text{for } s, t \geq 1.$$

Consider the change of variables to $(Z, V)$ defined by

$$S^{-1} = Z^\eta \cos^{2\alpha} V \quad \text{and} \quad T^{-1} = Z^\eta \sin^{2\alpha} V.$$
Chapter 7. Simulation methods

Then the density of \((Z, V)\) is given by

\[
f_{Z,V}(z,v) = \frac{\eta - \alpha}{\eta(2 - 2\alpha/\eta)} 2\sin v \cos v,
\]
defined within the domain

\[
\{(z, v) : 0 < z \leq \cos^{-2\alpha/\eta} v \text{ if } 0 < v \leq \pi/4 \text{ and } 0 < z \leq \sin^{-2\alpha/\eta} v \text{ if } \pi/4 < v < \pi/2\}.
\]

Noting that \(V\) has distribution function

\[
F_V(v) = \begin{cases} 
\frac{1 - \cos^{-2\alpha/\eta} v}{2 - 2\alpha/\eta} & \text{if } 0 < v \leq \pi/4, \\
\frac{1 - \cos^{-2\alpha/\eta} + \sin^{-2\alpha/\eta} v}{2 - 2\alpha/\eta} & \text{if } \pi/4 < v < \pi/2,
\end{cases}
\]

and that the conditional variable \(Z \mid V = v\) is uniformly distributed with density function

\[
f_{Z\mid V=v}(z) = \begin{cases} 
\cos^{2\alpha/\eta} v & \text{if } 0 < v \leq \pi/4, \\
\sin^{2\alpha/\eta} v & \text{if } \pi/4 < v < \pi/2,
\end{cases}
\]

we can represent \(Z\) and \(V\) in the following way:

\[
V = \begin{cases} 
\arccos \left[1 - (2 - 2\alpha/\eta)U_1 \right]^{\frac{\eta}{2(\eta - \alpha)}} & \text{if } 0 < U_1 \leq 1/2, \\
\arcsin \left[(2 - 2\alpha/\eta)U_1 + 2\alpha/\eta - 1 \right]^{\frac{\eta}{2(\eta - \alpha)}} & \text{if } 1/2 < U_1 < 1,
\end{cases}
\]

and

\[
Z = \begin{cases} 
U_2 \cos^{-2\alpha/\eta} V & \text{if } 0 < V \leq \pi/4, \\
U_2 \sin^{-2\alpha/\eta} V & \text{if } \pi/4 < V < \pi/2,
\end{cases}
\]

where \(U_1\) and \(U_2\) are independent variables with uniform distribution on \([0, 1]\).

Thus in order to generate iid observations from \(F_{ST}\) we follow the following construction:

- Generate independently \(U_1\) and \(U_2\) uniformly over \([0, 1]\).
- Define \(Z\) and \(V\) using equations (7.2) and (7.1).
- Set \(S = Z^{-\eta} \cos^{-2\alpha} V\) and \(T = Z^{-\eta} \sin^{-2\alpha} V\).

This is similar to the Shi et al. (1992) method for the logistic BEV case, although the construction is more intricate here.

Examples of points generated from \(F_{ST}\) for fixed \(\eta = 0.7\) and several values for \(\alpha\) are shown in Figure 7.1. This figure also contains the densities \(f_{ST}\) superimposed and is on a log-scale.
Figure 7.1: Simulated points from the bivariate logistic $F_{ST}$, with densities superimposed, using $\eta = 0.7$ and several values of $\alpha$. 

\begin{align*} 
\alpha &= 0.1 \\
\alpha &= 0.4 \\
\alpha &= 0.70001 \\
\alpha &= 1 \\
\alpha &= 1.4 \\
\alpha &= 2 \\
\alpha &= 3 \\
\alpha &= 10 \\
\alpha &= 100
\end{align*}
7.1.2 Using the modified asymmetric logistic model

Similar to the previous case, we consider the density function of the modified asymmetric logistic distribution function \( F_{ST} \) given in equation (4.18). This is given by

\[
f_{ST}(s, t) = \frac{\eta - \alpha}{\alpha \eta^2 q N_e} \left\{ s^{-1/\alpha} + \left( \frac{t}{q} \right)^{-1/\alpha} \right\}^{\alpha/\eta - 2} \left\{ s \left( \frac{t}{q} \right) \right\}^{-(1+1/\alpha)} \text{ for } s, t \geq 1.
\]

Now using the change of variables to \((Z, V)\) defined by \( S^{-1} = Z^\eta \cos^{2\alpha} V \) and \( qT^{-1} = Z^\eta \sin^{2\alpha} V \), the density of \((Z, V)\) is given by

\[
f_{ZV}(z, v) = \frac{\eta - \alpha}{\eta N_e} 2 \sin v \cos v,
\]

with domain

\[
\{(z, v) : 0 < z \leq \cos^{-2\alpha/\eta} v \text{ if } 0 < v \leq v^* \text{ and } 0 < z \leq \eta \cos^{1/\eta} \sin^{-2\alpha/\eta} v \text{ if } v^* < v < \pi/2\},
\]

where \( v^* = \arctan \eta^{1/(2\alpha)} \). Proceeding as above, it is easy to see that \( Z \) and \( V \) may be represented as

\[
V = \begin{cases} 
\arccos \left( \frac{1 - N_e U_1}{N_e^{\frac{2}{\alpha}}} \right) & \text{if } 0 < V \leq v^*, \\
\arcsin \left( \frac{\eta U_1 + (1 + \eta^{1/\alpha}) \frac{\pi}{\alpha} - 1}{N_e^{\frac{2}{\alpha}}} \right) & \text{if } v^* < V < \pi/2,
\end{cases}
\]

and

\[
Z = \begin{cases} 
U_2 \cos^{-2\alpha/\eta + 2} V & \text{if } 0 < V \leq v^*, \\
U_2 \eta \cos^{1/\eta} \sin^{-2\alpha/\eta + 2} V & \text{if } v^* < V < \pi/2,
\end{cases}
\]

where \( U_1 \) and \( U_2 \) are independent variables with uniform distribution on \([0,1]\).

Thus in order to generate iid observations from \( F_{ST} \) we follow the following construction:

- Generate independently \( U_1 \) and \( U_2 \) uniformly over \([0,1]\).
- Define \( Z \) and \( V \) using equations (7.4) and (7.3).
- Set \( S = Z^{-\eta} \cos^{2\alpha} V \) and \( T = q Z^{-\eta} \sin^{-2\alpha} V \).

Figure 7.2 depicts examples of points generated from the asymmetric logistic distribution function \( F_{ST} \) for fixed \( \alpha = 0.5 \) and \( \eta = 0.7 \) and several values of \( q \). This figure clearly shows the asymmetry caused by the dependence parameter \( q \) with the densities \( f_{ST} \) superimposed.
Figure 7.2: Simulated points from the bivariate asymmetric logistic $F_{ST}$, with densities superimposed, using $\alpha = 0.5$, $\eta = 0.7$ and a) $\varrho = 0.1$, b) $\varrho = 0.4$, c) $\varrho = 4$ and d) $\varrho = 10$. 
7.2 Simulation from the limiting processes

As mentioned above, Shi et al. (1992) suggested a scheme to generate from the standard BEV (symmetric) logistic distribution function. Methodologies to simulate from the standard multivariate symmetric and asymmetric logistic distribution functions are also developed by Stephenson (2002). We use here similar techniques for simulating from the symmetric and the asymmetric bivariate logistic limiting distribution function $G_\eta$.

7.2.1 Simulation from the bivariate logistic distribution function $G_\eta$

Consider the bivariate random variable $(X, Y)$ with joint distribution function $G_\eta$ with modified logistic dependence structure as in Example A, i.e. with the form

$$G_\eta(x, y) = \exp\left\{ -\left( 2 - 2^{\alpha/\eta} \right)^{-1} \left( x^{-1/\alpha} + y^{-1/\alpha} \right)^{\alpha/\eta} \right\},$$

(7.5)

if $\alpha \leq \eta$, for $x, y > 0$. Let $f_\eta$ denote the density function associated with $G_\eta$ and given by

$$f_\eta(x, y) = \eta^{-2} \left( 2 - 2^{\alpha/\eta} \right)^{-1} \exp\left\{ -\left( 2 - 2^{\alpha/\eta} \right)^{-1} \left( x^{-1/\alpha} + y^{-1/\alpha} \right)^{\alpha/\eta} \right\} (xy)^{(1+1/\alpha)} \times \left( x^{-1/\alpha} + y^{-1/\alpha} \right)^{\alpha/\eta-2} \left\{ \left( 2 - 2^{\alpha/\eta} \right)^{-1} \left( x^{-1/\alpha} + y^{-1/\alpha} \right)^{\alpha/\eta} + (\eta - \alpha)/\alpha \right\}.$$

Consider the transformation defined by

$$X^{-1} = \left\{ 2 - 2^{\alpha/\eta} \right\}^{-1} \eta \cos^{2\alpha} V \quad \text{and} \quad Y^{-1} = \left\{ 2 - 2^{\alpha/\eta} \right\}^{-1} \eta \sin^{2\alpha} V.$$

Straightforward algebraic manipulations give that the density of the bivariate random variable $(Z, V)$ is given by

$$f_{ZV}(z, v) = \exp(-z) \left\{ \frac{\alpha}{\eta} z + \left( 1 - \frac{\alpha}{\eta} \right) \right\} \sin 2v,$$

for $z > 0$ and $0 < v < \pi/2$.

From this representation it is clear that $Z$ and $V$ are independent with easily characterised distributions: $V$ may be represented as $\text{arcsin}(U^{1/2})$ where $U$ is uniform on $[0,1]$, while $Z$ is a $\left( 1 - \frac{\alpha}{\eta} \right) : \frac{\alpha}{\eta}$ mixture of a unit exponential random variable and the sum of two independent unit exponential random variables.

Alternatively, if we marginally transform from $(X, Y)$ to $(Z_1, Z_2)$ where the variables $\{Z_i \mid i = 1, 2\}$ are unit Fréchet distributed, that is, using the transformation...
Chapter 7. Simulation methods

Z_1 = (2 - 2^{\alpha/\eta}) X^{1/\eta} and Z_2 = (2 - 2^{\alpha/\eta}) Y^{1/\eta}, then the joint distribution function of (Z_1, Z_2) is given by

\[ G(z_1, z_2) = \exp \left\{ - \left( z_1^{-\alpha/\eta} + z_2^{-\alpha/\eta} \right)^{\alpha/\eta} \right\}, \]

for \( \alpha \leq \eta \), and \( z_1, z_2 > 0 \). This is the standard BEV logistic distribution function with dependence parameter \( 0 < \alpha' = \alpha/\eta \leq 1 \) and so points from this distribution can be obtained by a simple variant of the Shi et al. (1992) representation.

Figure 7.3 shows examples of points generated from \( G_\eta \) as in equation (7.5) for fixed \( \eta = 0.7 \) and several values of \( \alpha \), with densities superimposed.

7.2.2 Simulation from the bivariate asymmetric logistic distribution function \( G_\eta \)

We now present a methodology to simulate points from the limiting bivariate asymmetric logistic distribution function \( G_\eta \) which has the form

\[ G_\eta(x, y) = \exp \left( -N_{\theta \phi}^{-1} \left\{ \left( \frac{x}{\theta} \right)^{-1/\alpha} + \left( \frac{y}{\phi} \right)^{-1/\alpha} \right\}^{\alpha/\eta} + \frac{1 - \theta^{1/\eta}}{x^{1/\eta}} + \frac{1 - \phi^{1/\eta}}{y^{1/\eta}} \right) \]

(7.6)

if \( \alpha \leq \eta \), for \( x, y > 0 \). For this we first need to consider the following result which is similar to that given in Theorem 1 of Stephenson (2002).

Let \( (Z_1, V_1) \) be a bivariate random variable with joint distribution function \( G'_\eta \) of \( Z_1, V_1 \) given by

\[ G'_\eta(z_1, v_1) = \exp \left\{ -N_{\theta \phi}^{-1} \left( z_1^{-1/\alpha} + v_1^{-1/\alpha} \right)^{\alpha/\eta} \right\} \]

for \( z_1, v_1 > 0 \) and consider the two iid univariate random variables \( Z_2 \) and \( V_2 \) with distribution function \( F_Z \) given by \( F_Z(x) = \exp \left( -N_{\theta \phi}^{-1} x^{-1/\eta} \right) \) for \( x > 0 \). Then

\[
\Pr \left[ \max \left\{ \frac{\theta Z_1}{\phi}, \left( 1 - \theta^{1/\eta} \right)^{\eta} Z_2 \right\} \leq x, \max \left\{ \phi V_1, \left( 1 - \phi^{1/\eta} \right)^{\eta} V_2 \right\} \leq y \right] = \\
= \Pr \left( Z_1 \leq x/\theta, V_1 \leq y/\phi \right) \Pr \left( Z_2 \leq \left( 1 - \theta^{1/\eta} \right)^{-\eta} x \right) \Pr \left( V_2 \leq \left( 1 - \phi^{1/\eta} \right)^{-\eta} y \right) = \\
= \exp \left[ -N_{\theta \phi}^{-1} \left\{ \left( \frac{x}{\theta} \right)^{-1/\alpha} + \left( \frac{y}{\phi} \right)^{-1/\alpha} \right\}^{\alpha/\eta} \right] \exp \left( -N_{\theta \phi}^{-1} \frac{1 - \theta^{1/\eta}}{x^{1/\eta}} \right) \exp \left( -N_{\theta \phi}^{-1} \frac{1 - \phi^{1/\eta}}{y^{1/\eta}} \right) = G_\eta(x, y). 
\]

To exploit the above we note that \( G_\eta' \) has the form of a bivariate symmetric logistic distribution function with a normalising constant \( N_{\theta \phi}^{-1} \) instead of \( (2 - 2^{\alpha/\eta})^{-1} \). Consequently, it is sufficient to generate \( (Z_1, V_1) \) from the distribution function \( G_\eta' \), using
Figure 7.3: Simulated points from the bivariate logistic distribution function $G_\eta$, with densities superimposed, using $\eta = 0.7$ and several values of $\alpha$. 
the methodology given above but with \((2 - 2^{\alpha/\eta})\) replaced by \(N_{\theta \phi}\), and to generate the variables \(Z_2\) and \(V_2\) from the univariate Fréchet distribution \(F_Z\). The bivariate random variable

\[
\left( \max \left\{ \theta Z_1, \left( 1 - \theta^{1/\eta} \right)^{\eta} Z_2 \right\}, \max \left\{ \phi V_1, \left( 1 - \phi^{1/\eta} \right)^{\eta} V_2 \right\} \right)
\]

then has distribution function \(G_{\eta}\).

Again, an alternative method may be obtained by performing suitable transformations and using standard BEV results. Specifically, taking \(Z_1 = N_{\theta \phi} X^{1/\eta}\) and \(Z_2 = N_{\theta \phi} Y^{1/\eta}\) we have that \(Z_1\) and \(Z_2\) are unit Fréchet random variables with joint distribution function

\[
G(z_1, z_2) = \exp \left( - \left[ \left( \frac{z_1}{\theta^{1/\eta}} \right)^{-\eta/\alpha} + \left( \frac{z_2}{\phi^{1/\eta}} \right)^{-\eta/\alpha} \right]^{\alpha/\eta} + \frac{1 - \theta^{1/\eta}}{z_1} + \frac{1 - \phi^{1/\eta}}{z_2} \right),
\]

for \(\alpha \leq \eta\) and \(z_1, z_2 > 0\). This is the standard BEV asymmetric logistic distribution function with dependence parameters \(\alpha' = \alpha/\eta \in (0, 1]\), \(\theta' = \theta^{1/\eta} \in [0, 1]\) and \(\phi' = \phi^{1/\eta} \in [0, 1]\). Points from this latter distribution can be generated using the methodology of Stephenson (2002).

Examples of points generated from the bivariate asymmetric logistic distribution function \(G_{\eta}\) for fixed \(\alpha = 0.3\) and \(\eta = 0.7\) and several values of \(\theta\) and \(\phi\) are shown in Figure 7.4 with the corresponding densities superimposed.

### 7.3 Comments

In this chapter we developed methods for simulating from the joint tail model \(F_{ST}\) and also methods for simulating from the limiting distribution of the normalised componentwise maxima for the asymptotically independent case when using the logistic and asymptotic logistic dependence structures. Although simulating from the limiting distribution \(G_{\eta}\) turns out to be a special case of simulating from the standard BEV distribution function, our methods for simulating from the joint tail model \(F_{ST}\) are not of this type.
Figure 7.4: Simulated points from the bivariate asymmetric logistic distribution function $G_\eta$, with densities superimposed, using $\alpha = 0.3, \eta = 0.7$ and a) $\theta = 0.1$ and $\phi = 0.1$, b) $\theta = 0.1$ and $\phi = 0.9$, c) $\theta = 0.9$ and $\phi = 0.1$ and d) $\theta = 1$ and $\phi = 1$. 
Chapter 8

Extensions and further work

In Chapter 4 we derived a pseudo-polar representation of asymptotic independence for the bivariate case in terms of an angular measure density $h_\eta$ satisfying only the normalising condition (4.10). Parametric joint tail models to estimate the joint tail of the survivor function of a given bivariate random variable were thereafter obtained. We concentrate here on examining further properties of our models and on generalising these results to the multivariate case. First, a univariate model for partially observed data based on this framework is developed and then a generalisation of the pseudo-polar representation of asymptotic independence to the multivariate case is considered. Finally, some future areas of research are suggested.

8.1 Modelling dependence with partially observed data

Consider the bivariate random variable $(X, Y)$ with unit Fréchet margins and joint distribution function $F_{XY}$ satisfying equation (4.1). Our focus here will be modelling the joint structure of $(X, Y)$ when only the $X$ observations that correspond to $Y$ being large are observed. The conditional distribution function of $X$ given that $Y$ exceeds a high threshold $u$ has the form

$$F_{X|Y>u}(x) = \Pr(X \leq x \mid Y > u) = 1 - \frac{F_{XY}(x, u)}{F(u)}$$  \hspace{1cm} (8.1)$$

where $F$ is the unit Fréchet survivor function. Therefore the joint tail parametric models derived in Section 4.3 provide parametric models for the tail of the conditional
distribution function defined in equation (8.1). That is, we obtain a model for the tail of the univariate variable \( Z = X | Y > u \).

This model can be useful in cases where values of the \( X \) variable are measured or observed only when a corresponding variable \( Y \) takes values above some high threshold. Alternatively, the resulting \( X \) observations may be viewed as a non-random thinning of an underlying iid bivariate data set. Using only these partially observed data we are able to model the extremal dependence between the variables \( X \) and \( Y \) and undertake estimation using maximum likelihood, for example. The performance of this model is examined in the following example using the logistic dependence structure described in Chapter 4 and the simulated data studied in Chapter 5. We note that the procedure here is closely related to that of estimating \( r_j \) using the diagnostic method based on the structure variable \( V = \min(X, Y) \), as in Section 5.3.4. However, now we are able to identify more of the underlying model structure.

### 8.1.1 Example

For the logistic joint tail survivor model \( F_{XY} \) as in equation (4.14), the conditional distribution function in equation (8.1) has the form

\[
F_{X|Y>u}(x) = 1 - \frac{\lambda_u u^{1/\eta}}{2 - 2\alpha/\eta} \left\{ x^{-1/\eta} + u^{-1/\eta} - \left( x^{-1/\alpha} + u^{-1/\alpha} \right)^{\alpha/\eta} \right\}
\]

for \( x > u \), where \( \eta \in (0, 1], \alpha > 0, \lambda_u = P(X > u | Y > u) \) and \( u \) is a high threshold.

Consider the iid sequence of bivariate random variables \( \{(X_i, Y_i); i = 1, \ldots, n\} \) and define the univariate random variables \( Z_j = X_j | Y_j > u \) for \( j = 1, \ldots, n^* \), where \( 1, \ldots, n^* \) is an enumeration of the \( Y \) observations that exceed the threshold \( u \). Estimates of the dependence parameters \( \eta \) and \( \alpha \) and of the threshold exceedance probability \( \lambda_u \) can be obtained by the usual censoring-based likelihood methodology, i.e., for these \( n^* \) points, maximising the likelihood given by

\[
L_{n^*}(\eta, \alpha, \lambda_u; \mathbf{x}) = (1 - \lambda_u)^{n^* - n_u} \left\{ \frac{\lambda_u u^{1/\eta}}{\eta (2 - 2\alpha/\eta)} \right\}^{n_u} \times \prod_{j: z_j > u} \left\{ z_j^{-1+1/\eta} - \left( z_j^{-1/\alpha} + u^{-1/\alpha} \right)^{\alpha/\eta - 1} z_j^{-1+1/\alpha} \right\}
\]

where \( n_u \) is the number of points that exceeded the threshold \( u \). We assess the performance of the model in equation (8.2) by using the simulated bivariate data sets \( A, \ldots, 138 \)
Chapter 8. Extensions and further work

B, C and D of points \( \{(X_i, Y_i); i = 1, \ldots, 25,000\} \) given in Chapter 5 corresponding respectively to the BEV logistic, \( \rho = 0.5 \) bivariate normal, Morgenstern and \( \rho = -0.5 \) bivariate normal dependence structures. Then, we construct the univariate random variables \( Z_j = X_j \mid Y_j > u \) for \( j = 1, \ldots, n^* \) where \( u \) is the same threshold as was used in Section 5.4 for each data set.

Fitting the model (8.2) to the univariate data \( Z_j \) requires estimating the parameters \( \lambda_u, \eta \) and \( \alpha \). As discussed in Chapter 5, the threshold exceedance probability \( \lambda_u \) was estimated by its empirical analogue. Estimates of the dependence parameters \( \eta \) and \( \alpha \), obtained by maximising the likelihood function in equation (8.3), with associated standard errors based on the delta-method (in parentheses) are given Table 8.1 for each data set and corresponding threshold \( u \). For comparison with these values, the dependence parameter estimates obtained in Chapter 5 by fitting the modified logistic joint tail model to the bivariate data \( \{(X_i, Y_i); i = 1, \ldots, 25,000\} \) are also included in the table.

<table>
<thead>
<tr>
<th>Data set</th>
<th>( \eta )</th>
<th>( \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True Univ. Model Joint Model</td>
<td>Univ. Model Joint Model</td>
</tr>
<tr>
<td>A</td>
<td>1 0.95 (0.05) 0.99 (0.03)</td>
<td>0.86 (0.10) 0.76 (0.02)</td>
</tr>
<tr>
<td>B</td>
<td>0.75 0.73 (0.04) 0.75 (0.02)</td>
<td>0.99 (0.09) 0.97 (0.03)</td>
</tr>
<tr>
<td>C</td>
<td>0.5 0.58 (0.03) 0.53 (0.01)</td>
<td>0.96 (0.08) 1.08 (0.05)</td>
</tr>
<tr>
<td>D</td>
<td>0.25 0.36 (0.02) 0.34 (0.01)</td>
<td>0.69 (0.06) 0.72 (0.03)</td>
</tr>
</tbody>
</table>

Table 8.1: Dependence parameter estimates using the chosen threshold \( u \) for each data set. Estimates are obtained by fitting the logistic univariate model in equation (8.2) and the logistic joint tail model in equation (4.14). True values for \( \eta \) are also included and standard errors are given in parentheses.

From analysing the table we conclude that, although model (8.2) uses only univariate data consisting of those values of \( X \) for which \( Y \) has values above the threshold \( u \), it produces reasonable dependence parameter estimates. Indeed, these estimates are similar to those obtained by the joint tail model which uses all the data. We conclude here that given only the univariate observations \( X \) for which \( Y \) exceeds a high threshold, then reasonable estimates of the joint extremal dependence structure can be obtained.
8.2 Modelling dependence within multivariate joint tails

In this section, we concentrate on generalising the pseudo-polar representation of asymptotic independence given in Section 4.2 to the multivariate case.

Consider the d-dimensional random variable $X = (X_1, \ldots, X_d)$ with unit Fréchet margins and satisfying

$$\Pr(X_1 > x_1, \ldots, X_d > x_d) = F_X(x) = \mathcal{L}(x) \left( \prod_{i=1}^{d} x_i \right)^{-1/(dn)} \quad (8.4)$$

where $\mathcal{L}$ is a multivariate slowly varying function (see Bingham et al., 1987) and $\eta \in (0,1]$ is the coefficient of tail dependence. We remark that, subject to some regularity, all multivariate joint upper tails have a form of this type. However, this representation, which is based on d-dimensional regular variation, does not provide a complete tail characterisation of d-dimensional survivor functions since regular variation in d-dimensions does not characterise behaviour in sub-dimensional spaces. Let $g$ denote the limit function of $\mathcal{L}$, so that for all $x \in \mathbb{R}^d_+$ and $c > 0$

$$g(x) = \lim_{r \to \infty} \frac{\mathcal{L}(rx)}{\mathcal{L}(r1)} \quad \text{and} \quad g(cx) = g(x) \quad (8.5)$$

where 1 represents the d-dimensional vector $(1, \ldots, 1)$ and let $u$ denote a high threshold. As for the bivariate case, $g$ is also called the ray dependence function and it can be written as $g(x) = g_\ast (x_1/\Sigma x_1, \ldots, x_d/\Sigma x_d)$. We consider the behaviour of the d-dimensional random variable $S = (S_1, \ldots, S_d)$ defined by the following

$$F_S(s) = \Pr(S_1 > s_1, \ldots, S_d > s_d)$$

$$\quad = \lim_{u \to \infty} \Pr(X_1 > us_1, \ldots, X_d > us_d \mid X_1 > u, \ldots, X_d > u)$$

$$\quad = \lim_{u \to \infty} \frac{F_X(us)}{F_X(u1)} = g(s) \left( \prod_{i=1}^{d} s_i \right)^{-1/(dn)}$$

for all $s = (s_1, \ldots, s_d) \in [1, \infty)^d$ and $g$ as in equation (8.5).

Transforming to the pseudo-polar coordinates $R = \sum_{j=1}^{d} S_j$ and $W_j = S_j/R$ for $j = 1, \ldots, d$, defining $W = (W_1, \ldots, W_d)$ and assuming the density of $(R, W)$ exists then this density can be shown to satisfy

$$f(r, w) = r^{-(1+1/\eta)} h_\eta(w) \quad (8.6)$$
for $w \in S_d$ and $r \in \left[ \max \{1/w_j\}, \infty \right]$, where the function $h_\eta$ is a non-negative angular measure density on the $(d-1)$-dimensional unit simplex $S_d$ that is determined by $g_\ast(w)$ and $\eta$.

As was our strategy previously, we reconstruct the survivor function $\overline{F}_s$ from the density in equation (8.6). Letting $r^* = \max_{1 \leq j \leq d} (s_j/w_j)$, we have

$$\overline{F}_s(s) = \int_{S_d} \int_{r^*}^\infty r^{-(1+1/\eta)} h_\eta(w) \, dr \, dw \quad \text{(8.7)}$$

for $s \in [1, \infty)^d$. Therefore, providing $h_\eta$ is known, this representation can be used to derive parametric models for the joint survivor function $\overline{F}_s$ and hence models for $g_\ast$ and thus also for the joint tail of the survivor function $\overline{F}_x$.

The $d$-dimensional analogue of the normalising condition given in equation (4.10) is obtained by writing $s_1 = \cdots = s_d = k$ in equation (8.7). Following this approach, we obtain

$$\eta^{-1} = \int_{S_d} \left\{ \min_{1 \leq j \leq d} (w_j) \right\}^{1/\eta} h_\eta(w) \, dw.$$

Theorem 4.1 in Chapter 4 extends naturally to the following result:

**Theorem 8.1** Let $X = (X_1, \ldots, X_d)$ satisfy equation (8.4) so that

$$\lim_{u \to \infty} \Pr(X > ux \mid X > u1) = g_\ast \left( \frac{x_1}{\Sigma x_i}, \ldots, \frac{x_d}{\Sigma x_i} \right) \times \left( \prod_{i=1}^d x_i \right)^{-1/(dn)} \quad \text{(8.8)}$$

for $x = (x_1, \ldots, x_d) \in [1, \infty)^d$, where $g_\ast$ is the limit function defined following equation (8.5). Then, for $w_j = x_j/\Sigma x_i$, $j = 1, \ldots, d$, the limit function $g_\ast(w)$ satisfies

$$g_\ast(w) = \eta \left( \prod_{i=1}^d w_i \right)^{1/(dn)} \times \int_{S_d} \left\{ \min_{1 \leq j \leq d} \left( \frac{t_j}{w_j} \right) \right\}^{1/\eta} h_\eta(t) \, dt \quad \text{(8.9)}$$

where $t = (t_1, \ldots, t_d)$ and $h_\eta$ is a non-negative measure density on the $(d-1)$-dimensional unit simplex $S_d$ satisfying

$$\eta^{-1} = \int_{S_d} \left\{ \min_{1 \leq j \leq d} (w_j) \right\}^{1/\eta} h_\eta(w) \, dw. \quad \text{(8.10)}$$

Conversely, given any $h_\eta$ satisfying equation (8.10), then equations (8.8) and (8.9) define a valid joint survivor function.
Chapter 8. Extensions and further work

Proof. Follows similarly to the proof of Theorem 4.1.

We conclude that any non-negative measure density $h_\eta$ on $S_d$ satisfying the normalising condition (8.10) provides a joint survivor function for $S$, and thus yields a valid model for the joint tail of $F_X$. The joint tail model for the original variable $X$ can then be obtained using the approximation

$$F_X(x) = \lambda F_S(x/u)$$

for $x \in (u, \infty)^d$, where $\lambda$ is the joint threshold exceedance probability and $u$ is a high threshold.

Analogous point process theory can also be developed from the above, and the analogue of the Coles and Tawn (1991) result that relates the measure density $h_\eta$ and the dependence function $V_\eta$, as given in Section 6.2.1, extends naturally to the $d$-dimensional case.

8.2.1 Example

Examples of multivariate joint tail parametric models can be obtained as in Section 4.3 by modifying the dependence structure of the parametric families of MEV distributions described in Section 1.4.4. However, the task of verifying that the resulting measure density $h_\eta$ satisfies the normalising condition (8.10) is not always straightforward, as multiple integration is involved.

We give here a trivariate joint tail model based on a modification of the mixed dependence structure. In this case, it is easy to show that the resulting density $h_\eta$, which is constant over the 2-dimensional unit simplex $S_3$, satisfies the normalising condition (8.10) that for the trivariate case has the form

$$\eta^{-1} = \int_0^{1/3} \int_{w_2}^{1-2w_2} w_2^{1/\eta} h_\eta(w_1, w_2) dw_1 dw_2 + \int_0^{1/3} \int_{w_1}^{1-2w_1} w_1^{1/\eta} h_\eta(w_1, w_2) dw_1 dw_2$$

$$+ \int_0^{1/3} \int_{1-2w_2}^{1-w_2} (1-w_1-w_2)^{1/\eta} h_\eta(w_1, w_2) dw_1 dw_2$$

$$+ \int_{1/3}^1 \int_{1/2-1/2w_2}^{1-w_2} (1-w_1-w_2)^{1/\eta} h_\eta(w_1, w_2) dw_1 dw_2.$$  (8.11)

As in the bivariate case, this normalising condition says nothing about the behaviour of the measure density $h_\eta$ at the boundaries of the unit simplex $S_3$. As discussed previously, such masses are ignored here but can be chosen arbitrarily.
The modified joint tail mixed model has measure density \( h_\eta \) defined by
\[
h_\eta(w_1, w_2) = 3^{1/\eta} \eta^{-3}(\eta + 1)(2\eta + 1)
\]
for \((w_1, w_2)\) in the interior of \( S_3 \), yielding by equation (8.7) the model
\[
\bar{F}_S(s_1, s_2, s_3) = 3^{1/\eta}(s_1 + s_2 + s_3)^{-1/\eta}
\]
for \( s_1, s_2, s_3 > 1 \), and associated limit function
\[
g_\ast(w_1, w_2) = 3^{1/\eta}\{w_1w_2(1 - w_1 - w_2)\}^{1/(3\eta)}.
\]
This provides the following joint survivor model for \( x \):
\[
\bar{F}_X(x_1, x_2, x_3) = \lambda (3u)^{1/\eta}(x_1 + x_2 + x_3)^{-1/\eta}
\]
for \( x_1, x_2, x_3 > u \), where \( u \) is a chosen high threshold and \( \lambda \) is the joint threshold exceedance probability.

### 8.3 Further work

For simplicity, the results obtained in this thesis for joint tail modelling assume a density measure \( h_\eta \). However, we should like to extend these results to a more general measure based framework involving \( dH_\eta(w) \).

Many aspects of the new asymptotically independent joint tail model need yet to be developed even in the bivariate case. For example, it was seen in Chapter 5 that the logistic ray dependence model in equation (4.13) is not flexible enough to accurately represent the negative association case provided by the \( \rho = -0.5 \) bivariate normal dependence structure, see Section 5.4.1. This shows that, for statistical applications, our joint tail parametric model presents some limitations and therefore a more flexible parametric model should be developed for cases of negative association. A possible step to overcome these difficulties could be to derive non-parametric joint tail models. Since non-parametric estimators for the coefficient of tail dependence \( \eta \) are available, e.g. in Peng (1999) and Draisma et al. (2001), this involves finding a non-parametric estimator for the measure density \( h_\eta \) such that the normalising condition (4.10) is satisfied.

In this thesis we developed a framework for the pseudo-polar coordinates \( r = x + y \) and \( w = x/(x + y) \). Corresponding results may be available for alternative coordinate...
specifications, e.g. standard trigonometric polar coordinates. A good starting point to then obtaining a non-parametric estimator for $h_\eta$ may be provided by the methodology developed by Einmahl et al. (1997, 2001) to construct non-parametric estimators for the polar angular measure (or spectral measure) of a standard BEV distribution. Clearly, we would want that any non-parametric estimator for $h_\eta$ should automatically satisfy the normalisation condition (4.10). Non-parametric methods for simultaneously estimating $\eta$ and $h_\eta$ would also be useful, as $\eta$ should not be assumed to be known exactly in the measure density estimation. More generally, non-parametric procedures for simultaneous estimation of marginal and dependence features are required.

In Chapter 5 it was seen that our threshold selection method tends to produce low thresholds for cases of weak or negative extremal association. Indeed, for cases where the unit Fréchet marginally distributed bivariate random variable $(X, Y)$ has weak extremal dependence, thresholds which are high on the $V = \min(X, Y)$ scale may be low on the marginal scale. Therefore, an alternative joint tail model for the joint survivor function $\overline{F}_{XY}$, such as

$$
\overline{F}_{XY}(x, y) = \mathcal{L}\left\{1/\overline{F}(x), 1/\overline{F}(y)\right\}\left\{\overline{F}(x)\overline{F}(y)\right\}^{1/(2\eta)}
$$

where $\overline{F}$ is the unit Fréchet survivor function, may be useful for these cases. Using similar procedures to those in Chapter 4, we have

$$
\Pr(S > s, T > t) = \lim_{u \to \infty} \Pr(X > us, Y > ut \mid X > u, Y > u)
= \lim_{u \to \infty} g\left\{1/\overline{F}(us), 1/\overline{F}(ut)\right\}\left\{\overline{F}(us)\overline{F}(ut)\right\}^{1/(2\eta)}\left\{\overline{F}(u)\right\}^{-1/\eta}
$$

and then, we obtain the following joint tail model for the tail of the joint survivor function of $(X, Y)$

$$
\overline{F}_{XY}(x, y) = \lambda \lambda_F^{-1/\eta} g\left\{1/\overline{F}(x), 1/\overline{F}(y)\right\}\left\{\overline{F}(x)\overline{F}(y)\right\}^{1/(2\eta)}
$$

(8.12)

for $x, y > u$, where $\lambda = \Pr(X > u, Y > u)$, $\lambda_F = \overline{F}(u)$ and $u$ is a high threshold. Note that if we approximate $\overline{F}(x)$ by its first order term then, for large $x$ and $y$, model (8.12) reduces to the model derived in Chapter 4.

The new joint tail survivor model is valid only within a joint tail region. Outside this region, no model is given for the dependence structure and consequently there is no joint tail model available. An asymptotically independent joint model for all extremes, i.e. for the whole region $\mathbb{R}_+^2 \setminus [0, u] \times [0, u]$, would therefore be an important
development for joint tail modelling. A possible approach to achieve this may be
to define the distribution function of a bivariate random variable with unit Fréchet
margins in terms of the distribution function of the bivariate variable \((S, T)\) that has
domain \([1, \infty) \times [1, \infty)\), as defined in Section 4.2. That is, given a bivariate random
variable \((A, B)\) with unit Fréchet margins and joint distribution function \(F_{AB}\), we
estimate the whole tail of the distribution function \(F_{AB}\) by the model

\[
F_{AB}(a, b) = F_{ST}[F_S^{-1}\{\exp(-1/a)\}, F_T^{-1}\{\exp(-1/b)\}]
\]

for \((a, b) \in \mathbb{R}_+^2 \setminus [0, u]^2\), where \(F_{ST}\), \(F_S\) and \(F_T\) are respectively the distribution function
and the marginal distributions of the variable \((S, T)\).

The extremal behaviour of Markov chains has been characterised by applications of ex­
treme value techniques to time series of data. Smith et al. (1997) develop a framework
based on the assumption that the limiting behaviour of the chain can be considered
exact over a fixed high threshold \(u\) above which the extreme structure between con­
secutive variables is determined by a BEV distribution, i.e. the transition distribution
of a Markov chain at extreme levels is assumed to be a BEV distribution. Bortot and
Tawn (1998) also model the behaviour of Markov chains over a high threshold using
the Ledford and Tawn (1996, 1997) model as the transition distribution and determine
threshold-dependent summaries of the clustering of extremes. Similar approaches to
those taken in these papers could be developed using our asymptotically independent
joint tail models as the transition distribution of the Markov chain. Replacing the BEV
distribution with our joint tail model would enable both asymptotically dependent and
asymptotically independent chains to be modelled and then the existing results would
be a special case of this framework. This approach could be useful for time series data
modelling, e.g. in financial data.

Ledford and Tawn (1998) showed that the standard BEV logistic parameter \(\alpha \in (0, 1]\)
can be interpreted as the probability of the limiting componentwise maxima not oc­
curring in a single observation. An interpretation for the dependence parameter \(\alpha > 0\)
of our new modified logistic joint tail model would also be of interest. In Chapter 5 a
diagnostic to estimate the dependence parameter \(\eta\) is given. A diagnostic to estimate
\(\alpha\) would also be useful as a preliminary step in exploring observed data. Further appli­
cation of the new model to real data with different features is required for investigating
possible weaknesses of the new methodology and motivating the development of new
Chapter 8. Extensions and further work

The tail model given in this chapter for the univariate data comprising values of \( X \) for which \( Y \) exceeds a high threshold needs also to be explored. Application areas and real data analysis situations where the implementation of this methodology could be useful need to be examined.

For the multivariate joint tail modelling discussed in this chapter, much work needs yet to be done. In particular, for the trivariate case additional examples of joint tail parametric models need to be developed, for example by modifying the dependence structure of parametric families of 3-dimensional MEV distributions, as was done for the bivariate case in Section 4.3. This requires finding the necessary normalising constant for the measure density \( h_r \) to satisfy the normalising condition (8.10), which becomes harder as the number of dimensions increases. It would be also of interest to investigate the relationship between the coefficient of tail dependence \( \eta \) for the joint tail of the trivariate random variable \( (X, Y, Z) \) and the coefficients of tail dependence \( \eta_1, \eta_2 \) and \( \eta_3 \) for the joint tail of the corresponding bivariate variables \( (X, Y) \), \( (X, Z) \) and \( (Y, Z) \). Similar methods to those used by Schlather and Tawn (2002) relating the extremal coefficients in \( d \)-dimensions and the smaller dimensional extremal coefficients could be a starting point. More generally, examples of multivariate joint tail parametric models need to be derived and developed together with associated statistical procedures for exploiting them within applied data analysis.
Bibliography


