



UNIVERSIDADE DA BEIRA INTERIOR
Ciências

Coeficientes de Dependência em Campos Aleatórios Max-Estáveis

Cecília Maria Fernandes Fonseca

Tese para obtenção do Grau de Doutor em
Matemática Aplicada
(3º ciclo de estudos)

Orientador: Prof. Doutora Luísa Maria Jota Pereira Amaral
Co-orientador: Prof. Doutora Ana Paula André Martins Fernandes

Covilhã, Dezembro de 2012

Ao António e ao Afonso.

Agradecimentos

Ao terminar este trabalho quero, antes de mais, dirigir o meu profundo agradecimento à Professora Doutora Luísa Amaral pelo empenho, dedicação e indispensável orientação em todas as etapas. À Professora Doutora Ana Paula Fernandes, co-orientadora deste trabalho, expresso a minha gratidão pela permanente disponibilidade, pelas muitas sugestões e pelo apoio. De ambas recebi muitos esclarecimentos, sempre prestados com o maior interesse.

O meu agradecimento muito especial à Professora Doutora Helena Ferreira com quem muito aprendi e de quem recebi muitas sugestões e críticas pertinentes. O seu espírito científico, a sua capacidade de trabalho e de entrega são contagiantes, sendo um incentivo para quem tem o prazer de com ela trabalhar. Obrigada pelo muito tempo que, generosamente, comigo despendeu.

Quero agradecer ao Centro de Matemática da Universidade da Beira Interior e à Fundação para a Ciência e Tecnologia, no âmbito do projeto PTDC/MAT/108575//2008, o apoio financeiro que me concedeu para participar em diferentes eventos científicos, que foram importantes para divulgar o trabalho realizado. Ao Instituto Politécnico da Guarda pelo apoio financeiro através da bolsa de doutoramento PRO-TEC - 2^a edição, cuja seriação foi da responsabilidade da Fundação para a Ciência e Tecnologia.

Aos meus amigos Miné, Quim e Teresa o meu muito obrigado, pela sua paciência e pelas muitas palavras de ânimo e incentivo nos momentos mais difíceis.

À minha família agradeço o carinho e toda a compreensão pelas minhas muitas ausências, mas em especial ao António e ao Afonso que souberam percebê-las demonstrando-o, muitas vezes, com gestos de afeto. Ao Fernando acresce agradecer o esclarecimento das dúvidas na utilização do Latex.

Resumo

A quantificação e caracterização da dependência entre extremos de campos aleatórios é de primordial importância para a modelação estatística. Nos campos aleatórios Gaussianos a função de correlação caracteriza a estrutura de dependência mas o pressuposto de Gaussianidade é inadequado quando se pretende modelar o comportamento da cauda de uma distribuição. Neste caso, os coeficientes de dependência entre extremos de campos aleatórios, já existentes na literatura, desempenham um papel importante.

Neste trabalho estudamos a dependência entre valores extremos de campos aleatórios max-estáveis através da introdução de novos coeficientes de dependência. Assim, tendo em conta a importância de avaliar a dependência em diferentes direções, quando o campo aleatório é anisotrópico, apresentamos uma abordagem multivariada e multidirecional da dependência extremal através de matrizes de coeficientes de dependência, deduzimos relações entre coeficientes e propomos um algoritmo para os determinar. Segue-se a introdução de um coeficiente que resume a dependência extremal entre os máximos das variáveis indexadas em dois conjuntos disjuntos de localizações, que designamos madograma generalizado, e o estudo das suas propriedades. Na sequência do interesse em quantificar o impacto da ocorrência de um acontecimento extremo numa dada localização a um conjunto de localizações vizinhas, introduzimos os índices de contágio e estabilidade e estudamos as suas propriedades. O índice de contágio permite avaliar em que medida a ocorrência de uma excedência de um nível elevado se pode propagar e contagiar um conjunto de localizações vizinhas. O índice de estabilidade permite analisar a oscilação de um campo num conjunto de localizações relativamente a uma localização específica. Para todos os coeficientes introduzidos propomos estimadores, demonstramos algumas das suas propriedades e aplicamo-los em amostras de dados simulados e de dados reais. Por fim, introduzimos um modelo espacial autoregressivo de máximos, baseado no modelo de Haslett (1979), provamos algumas das suas propriedades e avaliamos a dependência extremal com os coeficientes propostos.

Palavras-chave

Campo aleatório, coeficientes de dependência, distribuição multivariada de valores extremos, max-estabilidade, estacionariedade.

Abstract

Quantifying and characterizing dependence between extreme values of random fields has great importance for statistical modeling. In Gaussian random fields the correlation function characterizes the dependence structure, but the Gaussian requirement is unsuited when trying to model the behavior of the tail of a distribution. In this case, dependence coefficients between extremes of random fields, existing in the literature, have an important role.

In this work we study the dependence between extreme values of max-stable random fields through the introduction of new dependence coefficients. Thus, given the importance of evaluating dependence in different directions when a random field is anisotropic, we present a multivariate and multidirectional approach for extremal dependence by matrices of dependence coefficients, deduce relationships between the coefficients and propose an algorithm to compute them. It follows the introduction of a coefficient that summarizes extremal dependence between the maxima of variables indexed in two disjoint sets of locations, called generalized madograma, and the study of its properties. Following the interest in quantifying the impact that an extreme event in a given location has in a neighboring set of locations, we introduce the contagion and stability indices and study their properties. The contagion index enables the evaluation of the effect that an exceedance of a high level has on a surrounding set of locations. The stability index allows the analysis of the regional smoothness of a random field associated to a specific location. We propose estimators for all the introduced coefficients, prove some of their properties and apply them to simulated and real data. Finally, we introduce a spatial autoregressive maximum model, analogous to the model introduced in Haslett (1979), prove several of its properties and evaluate its spatial dependence with the proposed coefficients.

Keywords

Random field, dependence coefficients, multivariate extreme value distribution, max-stability, stationarity.

Conteúdo

Introdução	1
1 Resultados preliminares	5
1.1 Introdução	5
1.2 Teoria clássica de valores extremos	6
1.2.1 Sucessões unidimensionais	6
1.2.2 Sucessões d -dimensionais	10
1.2.3 Campos aleatórios	15
2 Matrizes de coeficientes de dependência	19
2.1 Introdução	19
2.2 Dependência local em eventos extremos	20
2.3 Coeficiente médio de dependência na cauda superior	24
2.4 Exemplos	28
2.5 Algoritmo para determinar os coef. de dependência	33
2.6 Campo aleatório M4	35
3 Madograma generalizado	41
3.1 Introdução	41
3.2 Definição e propriedades	42
3.3 Estimação	45
3.4 Estudo de simulação: alguns exemplos	49
3.5 Aplicação a dados de precipitação	53
4 Índices de contágio e estabilidade	57
4.1 Introdução	57
4.2 Índice de contágio	58
4.3 Índice de estabilidade	60
4.4 Exemplos	62
4.5 Estimação	66
4.6 Aplicação a dados de precipitação	69

CONTEÚDO

5 Um modelo autoregressivo de máximos	71
5.1 Introdução	71
5.2 Algumas fórmulas básicas	72
5.3 Estacionariedade	74
5.4 Max-estabilidade	75
5.5 Coeficientes de dependência	77
5.5.1 Matrizes de coeficientes de dependência na cauda superior . .	80
5.5.2 Madograma generalizado	81
5.5.3 Índices de contágio e estabilidade	83
Conclusões	87
Bibliografia	89

Lista de Figuras

1.1	Distribuições de valores extremos (Gumbel, Fréchet e Weibull) com $\alpha = 1$	8
2.1	Simulação do campo aleatório \mathbf{X} definido em (2.13) (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 114.3488$ (direita).	28
2.2	Representação, numa escala de cinzas, dos valores de dependência da matriz $\Lambda^3(\mathbf{i})$	30
2.3	Modelo de Schlather com $\rho(\ \mathbf{i} - \mathbf{j}\) = \exp(-\sqrt{(i_1 - j_1)^2 + (i_2 - j_2)^2})$	31
2.4	Dependência entre \mathbf{i} e $\mathcal{V}^n(\mathbf{i})$, $n \geq 1$, no modelo de Schlather.	32
2.5	Algoritmo para o cálculo do coeficiente de dependência multivariada na cauda superior	33
2.6	Simulação da função (2.18) com $\delta = 2$ (esquerda) e da função (2.19) com $p = (1/92, 1/95)$, $\delta_{12} = 1$ e $\theta = 2$ (direita).	35
2.7	Simulação do campo aleatório M4 (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 19.3613$ (direita).	38
2.8	Representação, numa escala de cinzas, dos valores de dependência da matriz $\Lambda^4(\mathbf{A})$	39
3.1	Simulação do campo aleatório M4 definido no Exemplo 3.1 (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 18.5626$ (direita).	50
3.2	Simulação do campo aleatório M4 definido no Exemplo 3.2 (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 20.2763$ (direita).	51
3.3	Média dos valores estimados do madograma generalizado $(\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J}))$ e erro quadrático médio (MSE) para o Exemplo 3.1	52
3.4	Valores exatos do madograma generalizado $\nu^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$ e média dos valores estimados $\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$ para o Exemplo 3.1.	52
3.5	Média dos valores estimados do madograma generalizado $(\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J}))$ e erro quadrático médio (MSE) para o Exemplo 3.2	52

LISTA DE FIGURAS

3.6	Valores exatos do madograma generalizado $\nu^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$ e média dos valores estimados $\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$ para o Exemplo 3.2.	53
3.7	Localização das estações, obtida a partir do SNIRH (esquerda), e a sua representação em coordenadas Lambert (direita).	54
3.8	Estimativas do madograma generalizado ($\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$) obtidas para $\mathbf{J} = \{\text{B.C.M.}\}$ e $\mathbf{I} = \{\text{L.Comprida, Fajão}\}$	54
3.9	Estimativas do madograma generalizado ($\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$) obtidas para $\mathbf{J} = \{\text{Penamacor}\}$ e $\mathbf{I} = \{\text{L.Comprida, Fajão}\}$	55
3.10	Estimativas do madograma generalizado ($\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$) obtidas para $\mathbf{J} = \{\text{C.Felgueira}\}$ e $\mathbf{I} = \{\text{L.Comprida, Fajão}\}$	55
3.11	Estimativas do madograma generalizado ($\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$) para $\alpha = \frac{\lambda}{2}$ e $\beta = 1 - \lambda$ com $\lambda \in (0, 1)$	55
3.12	Estimativas do madograma generalizado ($\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$) obtidas para $\mathbf{J} = \{\text{C.Felgueira}\}$, $\mathbf{I} = \{\text{L.Comprida, Fajão}\}$ com $\alpha = \frac{\lambda}{2}$ e $\beta = 1 - \lambda$, $\lambda \in (0, 1)$, e estimativas do λ -madograma para as restantes regiões.	56
4.1	Simulação do campo aleatório M4 definido no Exemplo 4.1 (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 16.4669$ (direita).	63
4.2	Simulação do campo aleatório M4 definido no Exemplo 4.2 (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 19.2996$ (direita).	64
4.3	Simulação do campo aleatório M4 definido no Exemplo 4.3 (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 19.5652$ (direita).	66
4.4	Localização das estações, obtida no SNIRH (esquerda), e a sua representação em coordenadas Lambert (direita).	69
5.1	Simulação do campo aleatório ARMAX com $k = 0.65$ (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 3.0174$ (direita).	72
5.2	Valores de $\varepsilon(\mathbf{i}, T_{s_3}^m(\mathbf{i}))$ para $m \in \{1, \dots, 25\}$ e $k \in \{0.01, 0.04m : m \in \{1, \dots, 24\}\}$	80
5.3	Representação, numa escala de cinzas, dos valores da matriz $\Lambda^3(\mathbf{i})$	81
5.4	Representação do madograma generalizado (5.14) para $\delta \in \{1, \dots, 25\}$ e $k \in \{0.01, 0.04m : m \in \{1, \dots, 24\}\}$	83
5.5	Representação do coeficiente extremal (5.16) para $\delta \in \{1, \dots, 25\}$ e $k \in \{0.01, 0.04m : m \in \{1, \dots, 24\}\}$	85

Lista de Tabelas

4.1	Média dos valores estimados do índice de contágio, \widehat{CI} , e erro quadrático médio, MSE , para o Exemplo 4.2 onde $CI(\mathbf{A}^{(k)}, \mathbf{i}) = 8$, k par.	67
4.2	Média dos valores estimados dos índices de contágio, \widehat{CI} , e de estabilidade, \widehat{SI} , e do erro quadrático médio, MSE , para o Exemplo 4.1 onde $CI(\mathbf{A}, \mathbf{i}) = 0$ e $SI(\mathbf{A}, \mathbf{i}) = 1$	68
4.3	Média dos valores estimados dos índices de contágio, \widehat{CI} , e de estabilidade, \widehat{SI} , e do erro quadrático médio, MSE , para o Exemplo 4.2 onde $CI(\mathbf{A}^{(k)}, \mathbf{i}) = 4.7$ e $SI(\mathbf{A}^{(k)}, \mathbf{i}) = 6$, k ímpar.	68
4.4	Média dos valores estimados dos índices de contágio, \widehat{CI} , e de estabilidade, \widehat{SI} , e do erro quadrático médio, MSE , para o Exemplo 4.3 onde $CI(\mathbf{A}, \mathbf{i}) \approx 4.5353$ e $SI(\mathbf{A}, \mathbf{i}) \approx 2.0712$	68
4.5	Estimativas dos índices de contágio e de estabilidade nas regiões {Gouveia, Oliveira do Hospital, Seia} e {Penamacor, Barragem Cabeço Monteiro} relativamente a "Lagoa Comprida".	69

Abreviaturas e Acrónimos

ARMAX	autoregressivo de máximos
BEV	bivariada de valores extremos
CI	índice de contágio
EVD	distribuição de valores extremos
f.d.	função de distribuição
GEV	generalizada de valores extremos
i.i.d.	independentes e identicamente distribuídas
MEV	multivariada de valores extremos
M4	máximos multivariados de máximos móveis
SI	índice de estabilidade
SNIRH	Sistema Nacional de Informação de Recursos Hídricos
v.a.	variável aleatória

Introdução

A teoria de valores extremos é um ramo da probabilidade e estatística que se dedica, essencialmente, ao estudo da ocorrência de acontecimentos extremos. O furacão Katrina em 2005, o crash do mercado de ações em 2007, as inundações no Paquistão em 2010, o furacão Irene em 2011 ou o maremoto no Japão em 2011 são alguns exemplos de acontecimentos extremos. Todos tiveram um impacto direto na vida das pessoas e na economia, acrescentando o efeito sobre os ecossistemas no caso particular dos eventos meteorológicos extremos. Estes fenómenos, relativamente recentes, revelam que apesar dos acontecimentos extremos terem uma pequena probabilidade de ocorrerem, quando se realizam, podem conduzir a consequências devastadoras, o que justifica a importância dada à teoria de valores extremos.

Das diversas áreas de aplicação da teoria de valores extremos, a da climatologia e da meteorologia têm sido alvo de interesse nos últimos anos (Schlather e Tawn (2003), Cooley (2005), Sang (2008) e Davison *et al.* (2012)), onde é conveniente estudar não só o comportamento médio das diferentes variáveis meteorológicas (precipitação, temperatura, velocidade do vento, entre outras), mas também dos seus valores extremos. A análise dos dados e os modelos climatéricos sugerem que eventos extremos associados a fenómenos meteorológicos irão aumentar no futuro (Davison *et al.* (2012)). As alterações no ciclo hidrológico, incluindo as cheias que se têm registado, vão de encontro a esta perspetiva (Trenberth (1998, 1999)). Assim, é de primordial importância, estudar o comportamento extremal de fenómenos meteorológicos, possibilitando a sua modelação, de modo a prever eventos extremos e, desta forma, definir estratégias de ação que acautelem pessoas e bens.

Muitos destes fenómenos aleatórios estão associados a variáveis meteorológicas definidas no tempo e no espaço, o que impulsionou o desenvolvimento da teoria de valores extremos em campos aleatórios sobre \mathbb{R}^d , $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{R}^d}$.

Apesar dos campos aleatórios Gaussianos possuírem boas propriedades, nomeadamente a estrutura de dependência ser caracterizada pela função de correlação, estes não se adaptam à modelação de valores extremos, surgindo assim a necessidade de encontrar formas de avaliar a dependência entre extremos espaciais.

A classe de campos aleatórios que, teoricamente, se justifica como sendo a mais adequada ao estudo de valores extremos no espaço, é a dos campos max-estáveis, as-

INTRODUÇÃO

sumindo um papel de destaque no estudo de casos práticos. Por exemplo, encontram-se aplicações destes processos na modelação da precipitação numa dada região em Coles (1993), Smith (1990), Coles e Tawn (1996), Schlather (2002) e Buishand *et al.* (2008).

Na modelação de acontecimentos extremos através de campos aleatórios max-estáveis a quantificação da sua dependência é inevitavelmente um tópico de grande importância que tem sido estudado por vários autores. Neste âmbito têm sido introduzidos alguns coeficientes de dependência extremal (Schlather e Tawn (2003), Cooley (2005), Cooley *et al.* (2006) Naveau *et al.* (2009)) sendo de destacar a importância que lhes tem sido atribuída pela comunidade científica, visível através da sua ampla utilização (Vannitsem e Naveau (2007), Bel *et al.* (2008), Sang (2008), Davison *et al.* (2012)).

Dos coeficientes de dependência extremal em campos aleatórios max-estáveis e estacionários realçamos a função coeficiente extremal apresentada em Schlather e Tawn (2003) e o F -madograma descrito em Cooley (2005). A função coeficiente extremal é uma adaptação ao contexto espacial do coeficiente extremal bivariado de Tiago de Oliveira (1962/63). Para campos max-estáveis e isotrópicos destacamos o λ -madograma introduzido por Naveau *et al.* (2009). Os dois últimos coeficientes, F -madograma e λ -madograma, são baseados no conceito de semivariograma utilizado em Geoestatística. Todos estes coeficientes permitem uma abordagem bivariada da dependência espacial.

Os coeficientes de dependência multivariada, existentes na literatura, também desempenham um papel importante na avaliação da dependência extremal entre vetores de variáveis de um campo aleatório max-estável \mathbf{X} . São exemplo disso o coeficiente extremal multivariado de Smith(1990) e o coeficiente de dependência na cauda multivariada (Li (2009), Ferreira (2011)) os quais são extensões, respetivamente, do coeficiente extremal bivariado de Tiago de Oliveira (1962/63) e do coeficiente de dependência na cauda superior de Sibuya (1960).

Com este trabalho pretendemos enriquecer o estudo da dependência extremal em campos aleatórios max-estáveis, propondo novos coeficientes que visam responder à necessidade de quantificar a dependência espacial em determinadas condições, que passaremos a descrever.

Após uma breve apresentação sobre conceitos e resultados relevantes da teoria de valores extremos essenciais ao enquadramento do trabalho desenvolvido (Capítulo 1), introduzimos no Capítulo 2 uma abordagem à dependência extremal de campos aleatórios max-estáveis $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{Z}^2}$ através de matrizes de coeficientes de dependência multivariada, num contexto multidirecional. A motivação desta proposta reside no facto de num campo aleatório anisotrópico ser necessário avaliar a dependência espacial em diferentes direções. Tal questão não se coloca se o campo

aleatório \mathbf{X} é isotrópico, onde os coeficientes de dependência extremal no contexto bivariado apenas dependem da distância entre as localizações das variáveis consideradas. Os elementos das matrizes são coeficientes de dependência multivariada na cauda superior ou coeficientes extremais no contexto multivariado. Estes coeficientes avaliam a dependência entre determinados acontecimentos extremos associados às variáveis de \mathbf{X} indexadas em dois conjuntos disjuntos e vizinhos de localizações. Dado que o cálculo dos coeficientes de dependência multivariada na cauda superior pode ser uma tarefa árdua, propomos um algoritmo para o efeito. Ainda no mesmo capítulo, deduzimos relações entre coeficientes de dependência multivariada e apresentamos um novo coeficiente definido pela média de coeficientes de dependência multivariada na cauda superior.

Os coeficientes de dependência extremal referidos anteriormente para campos aleatórios max-estáveis estacionários e/ou isotrópicos, baseiam-se no comportamento de pares de variáveis de um campo \mathbf{X} . Contudo, é por vezes conveniente avaliar a dependência entre os valores extremos de variáveis indexadas em dois conjuntos disjuntos de localizações, de forma a considerar a dependência intra-conjuntos e inter-conjuntos. A relevância deste aspecto é reconhecida pelos hidrologistas ao agruparem dados por regiões com base nas suas características hidrográficas e topográficas. O mesmo aspecto é reforçado pela influência da corrente marítima do Golfo do México na temperatura das ilhas Britânicas e pelo efeito que o sistema montanhoso português Sintra-Montejunto-Estrela tem na precipitação das regiões vizinhas. Assim, são necessárias medidas que possam quantificar a dependência entre os valores extremos de regiões disjuntas. Este facto motivou-nos a introduzir, no Capítulo 3, o madograma generalizado para campos max-estáveis $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{R}^2}$. Este novo coeficiente resume a dependência entre o máximo das variáveis de um campo \mathbf{X} localizadas em duas regiões disjuntas e generaliza o λ -madograma (Naveau *et al.* (2009)). Estabelecemos ainda relações entre este coeficiente e outros coeficientes de dependência que existem na literatura.

Quando ocorre um fenómeno meteorológico extremo numa dada localização é aconselhável avaliar o seu impacto em regiões vizinhas, por exemplo, a precipitação elevada numa dada localização poderá ter consequências graves em regiões contíguas. Este foi o motivo que esteve na base da introdução de dois novos coeficientes definidos num campo max-estável $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{R}^2}$, e que denominámos de índices de contágio e de estabilidade, aos quais dedicamos o Capítulo 4. O índice de contágio quantifica o efeito que a excedência de um valor elevado numa localização pode ter nas variáveis de um campo indexadas num conjunto de localizações vizinhas, permitindo avaliar a propagação desse evento. O índice de estabilidade reflete a oscilação de um campo aleatório num conjunto de localizações face a uma localização \mathbf{i} , com base no número esperado de cruzamentos de um nível elevado a partir de \mathbf{i} . Estu-

INTRODUÇÃO

dámos ainda as propriedades destes coeficientes e deduzimos a sua relação com o coeficiente extremal bivariado.

Nos Capítulos 3 e 4 apresentamos ainda estimadores não paramétricos para o madograma generalizado e para os índices de contágio e de estabilidade, respetivamente. Estudamos a sua consistência e o seu comportamento assintótico. Descrevemos os estudos de simulação que realizámos para avaliar o seu desempenho recorrendo a modelos cuja estrutura de dependência é conhecida. No estudo de simulação utilizámos um campo aleatório baseado nos processos máximos multivariados de máximos móveis (M4). Apresentamos também os resultados da aplicação dos estimadores em amostras de dados reais obtidas em <http://snirh.pt/>, o Sistema Nacional de Informação de Recursos Hídricos de Portugal (SNIRH). Neste sítio obtivemos dados de precipitação máxima anual, com base em máximos diários, registados através de estações meteorológicas em diferentes localizações de Portugal. Seleccionámos as localizações de Portugal Continental que considerámos terem interesse para ilustrar a aplicação dos coeficientes, tendo por base as suas características geográficas e climatéricas.

Com base no trabalho de Haslett (1979) propomos, no Capítulo 5, um campo aleatório autoregressivo de máximos. Estudam-se algumas propriedades deste modelo, nomeadamente, a estacionariedade e max-estabilidade. Neste campo aleatório avaliamos a dependência extremal através dos coeficientes que introduzimos nos Capítulos 2, 3 e 4.

Reservámos para as Conclusões um resumo dos resultados obtidos e a identificação de alguns tópicos que poderão ser objeto de trabalho futuro.

Salientamos o papel fundamental dos meios informáticos no desenvolvimento deste trabalho. Para aplicarmos os estimadores propostos em amostras de dados simulados e de dados reais, implementámo-los no programa de acesso livre *R* (<http://cran.r-project.org>). Também utilizámos o *Mathematica 7* para desenvolver o algoritmo de cálculo dos coeficientes de dependência multivariada na cauda. A utilização deste programa justifica-se pela sua capacidade de cálculo simbólico o que permitiu obter a expressão algébrica dos coeficientes. Recorremos ainda ao *Matlab* para obter a simulação de alguns campos aleatórios.

Capítulo 1

Resultados preliminares

1.1 Introdução

A teoria de valores extremos estuda o comportamento estocástico de extremos associados a um conjunto de variáveis aleatórias (v.a.'s) ou vetores aleatórios. A abordagem clássica consiste em estudar o comportamento assintótico da cauda superior e/ou inferior da função de distribuição (f.d.) comum, F , de um conjunto de variáveis aleatórias e trabalhar no sentido de o modelar e de desenvolver técnicas estatísticas que permitam estimar parâmetros inerentes a esse modelo.

São muitos os trabalhos nesta área, sendo uns mais centrados nos seus fundamentos matemáticos, de que são exemplo Galambos (1978), Leadbetter (1983), Resnick (1987, 2006), de Haan e Ferreira (2006), outros focam-se mais nas aplicações e na análise de dados como é o caso de Embrechts *et al.* (1997), Coles (2001) e Beirlant *et al.* (2004), entre outros.

A importância desta teoria está espelhada na diversidade de áreas que a utiliza, vejam-se os exemplos da sua aplicação apresentados por Beirlant *et al.* (2004), e que vão desde a avaliação de risco nas áreas financeira e de seguros, aos estudos de fiabilidade e de resistência de materiais, na engenharia. Contudo, onde tem sido mais aplicada é na meteorologia e hidrologia.

Tem sido crescente o desenvolvimento de ferramentas informáticas que permitem aplicar muitos dos resultados da teoria de valores extremos. Pensamos que a sua disponibilização gratuita propiciou a sua crescente utilização. Exemplo disso, é o vasto conjunto de *packages* disponíveis no programa *R*. Deste, destacamos para análise de extremos em sucessões univariadas as *packages ismev*, *evd* e *extremevalues* e, para análise de extremos em sucessões d -dimensionais e no contexto espacial, as *packages evd*, *RandomFields* e *SpatialExtremes*.

Neste capítulo introduzem-se os conceitos e resultados utilizados no decorrer deste trabalho, bem como a notação que será utilizada.

1.2 Breve referência à teoria de valores extremos

A evolução na teoria de valores extremos foi movida pela necessidade de descrever e encontrar regularidade no comportamento extremal de diferentes fenómenos aleatórios. Em 1943, surgiram os primeiros modelos matemáticos, assumindo a independência e a igualdade das distribuições de probabilidade das variáveis de uma sucessão. Contudo, à medida que cresceu a necessidade de resolver problemas práticos foi necessário enfraquecer as referidas condições, introduzindo-se a dependência entre as variáveis e/ou a não igualdade das distribuições das variáveis.

Refira-se, ainda, que inicialmente foram estudadas sucessões unidimensionais, mas a necessidade de resolver problemas multivariados propiciou o desenvolvimento da teoria de valores extremos em sucessões multidimensionais. Por outro lado, o estudo de extremos de processos naturais que variam no tempo e no espaço, por exemplo climatológicos, impulsionou o desenvolvimento da teoria de valores extremos em campos aleatórios.

1.2.1 Sucessões unidimensionais

Seja $\mathbf{X} = \{X_n\}_{n \geq 1}$ uma sucessão de v.a.'s independentes e identicamente distribuídas (i.i.d.) com f.d. comum, F .

A abordagem clássica da teoria de valores extremos consiste em caracterizar a cauda da distribuição F a partir da distribuição do máximo das n primeiras variáveis da sucessão \mathbf{X} , que definimos por

$$M_n = \bigvee_{i=1}^n X_i,$$

onde ' \bigvee ' denota o máximo. Sendo a função de distribuição de M_n dada por

$$P(M_n \leq x) = F^n(x), \quad x \in \mathbb{R}, \quad n \in \mathbb{N}, \quad (1.1)$$

então, para se conhecer a distribuição exata do máximo precisaríamos de especificar a distribuição F , a qual é, em geral, desconhecida. Surge assim o interesse em obter um resultado assintótico menos dependente de F . Ao considerarmos o limite da f.d. de M_n obtemos uma distribuição degenerada,

$$\lim_{n \rightarrow +\infty} F^n(x) = \begin{cases} 0 & \text{se } x < x_F \\ 1 & \text{se } x = x_F \end{cases}$$

onde

$$x_F = \sup\{x : F(x) < 1\},$$

denota o limite superior do suporte de F , sendo, portanto, um resultado desprovido de interesse. Este resultado aliado ao facto de a f.d. F ser, usualmente, desconhecida levantou as duas questões que passaremos a apresentar.

A primeira consistiu em averiguar se para M_n sujeito a normalização linear se teria como limite em (1.1) uma f.d. não degenerada, isto é, se

$$P(a_n^{-1}(M_n - b_n) \leq x) \xrightarrow[n \rightarrow +\infty]{w} G(x), \quad (1.2)$$

para alguma f.d. G não degenerada e sucessões de constantes reais $\{a_n > 0\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$, onde $\xrightarrow[n \rightarrow +\infty]{w}$ denota a convergência em todos os pontos de continuidade de G . Quando tal acontece dizemos que F pertence ao domínio de atração de G (para o máximo) e escrevemos $F \in D(G)$.

A resposta a esta questão constitui o resultado central da teoria de valores extremos, conhecido por Teorema dos Tipos Extremos, devido a Fréchet (1927), Fisher e Tippett (1928) e Gnedenko (1943), o qual estabelece as possíveis distribuições limite para o máximo de sucessões i.i.d., devidamente normalizado.

Teorema 1.1 (Teorema de Tipos Extremos) *Seja \mathbf{X} uma sucessão de v.a.'s i.i.d. com f.d. comum F . Se \mathbf{X} verifica (1.2) para algumas sucessões de constantes $\{a_n > 0\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ então G é do tipo de uma das seguintes distribuições:*

$$\text{Tipo I (Gumbel)} \quad H(x) \equiv \Lambda(x) = \exp(-\exp(-x)), \quad x \in \mathbb{R};$$

$$\text{Tipo II (Fréchet)} \quad H(x) \equiv \Phi_\alpha(x) = \exp(-x^{-\alpha}), \quad x > 0, \alpha > 0;$$

$$\text{Tipo III (Weibull)} \quad H(x) \equiv \Psi_\alpha(x) = \exp(-(-x)^{-\alpha}), \quad x \leq 0, \alpha > 0.$$

Quando dizemos que G é do tipo de uma das três distribuições dadas no Teorema 1.1, conhecidas como distribuições de valores extremos, queremos dizer que existem constantes reais $a > 0$ e b tais que $G(x) = H(ax + b)$.

A classe de leis limite, G , em (1.2) coincide com a classe das distribuições max-estáveis, isto é, distribuições H tais que

$$H^k(A_k x + B_k) = H(x), \quad x \in \mathbb{R},$$

para algumas constantes reais $A_k > 0$, B_k , com $k \geq 1$.

A segunda questão consistiu em caracterizar a classe de distribuições F para as quais (1.2) se verifica, ou seja, identificar o domínio de atração de G , para cada lei limite G . A correspondência entre a f.d. F e o tipo de distribuição limite é estabelecida a partir de condições sobre a cauda de F . Entre os trabalhos referentes

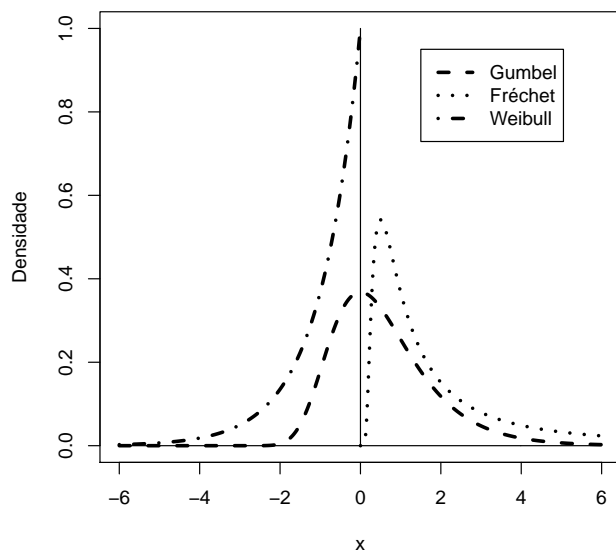


Figura 1.1: Distribuições de valores extremos (Gumbel, Fréchet e Weibull) com $\alpha = 1$.

à caracterização de cada um dos domínios de atração destacamos os que se encontram em von Mises (1954), Gnedenko (1943), de Haan (1970), entre outros.

As expressões analíticas das distribuições dadas no Teorema 1.1 podem ser unificadas na seguinte forma paramétrica, conhecida como forma de von Mises-Jenkinson ou função generalizada de valores extremos (GEV),

$$G_\gamma(x) = \begin{cases} \exp\left(-(1 + \gamma x)^{-1/\gamma}\right) & \text{se } \gamma \neq 0, 1 + \gamma x > 0 \\ \exp(-\exp(-x)) & \text{se } \gamma = 0 \end{cases}. \quad (1.3)$$

Para $\gamma < 0$, $\gamma = 0$ e $\gamma > 0$ a f.d. G_γ reduz-se às f.d.'s de Weibull, Gumbel e Fréchet, respetivamente (Figura 1.1). O parâmetro γ , designado por índice de cauda, determina o tipo de cauda de F , nomeadamente, uma cauda curta se $\gamma < 0$ (domínio de atração Weibull), uma cauda exponencial se $\gamma = 0$ (domínio de atração Gumbel) e uma cauda longa se $\gamma > 0$ (domínio de atração Fréchet).

Com a introdução dos parâmetros de localização, $\lambda \in \mathbb{R}$, e escala, $\delta > 0$, a função GEV é dada por

$$G_{\gamma,\lambda,\delta}(x) = \begin{cases} \exp\left(-\left(1 + \gamma\left(\frac{x-\lambda}{\delta}\right)\right)^{-1/\gamma}\right) & \text{se } \gamma \neq 0, 1 + \gamma\left(\frac{x-\lambda}{\delta}\right) > 0 \\ \exp\left(-\exp\left(-\frac{x-\lambda}{\delta}\right)\right) & \text{se } \gamma = 0 \end{cases}.$$

Depois de conhecidas as possíveis formas da distribuição limite do máximo de n v.a.'s i.i.d. houve necessidade de saber em que condições se tinha a convergência para este limite. O próximo resultado estabelece a condição necessária e suficiente

a impor a uma sucessão arbitrária de números reais $\{u_n\}_{n \geq 1}$, não necessariamente da forma $u_n = a_n x + b_n$, para que se tenha a convergência de $P(M_n \leq u_n)$.

Proposição 1.1 *Seja \mathbf{X} uma sucessão de v.a.'s i.i.d. com f.d. comum F e $\{u_n\}_{n \geq 1}$ uma sucessão de números reais. Então, para $0 \leq \tau < \infty$, tem-se*

$$P(M_n \leq u_n) \xrightarrow{n \rightarrow +\infty} \exp(-\tau)$$

se e somente se

$$n(1 - F(u_n)) \xrightarrow{n \rightarrow +\infty} \tau. \quad (1.4)$$

Quando (1.4) se verifica, dizemos que u_n é um nível normalizado e representamo-lo por $u_n^{(\tau)}$. Este resultado ilustra a dependência do comportamento da cauda da distribuição F e a existência de distribuição limite para o máximo.

O facto de em grande parte das situações reais a hipótese de independência não ser verificada fez com que a teoria de valores extremos passasse a contemplar sucessões dependentes. Sob estacionariedade e condições convenientes de dependência fraca, amplamente discutidas em Leadbetter (1974), Leadbetter *et al.* (1983), Leadbetter e Nandagopalan (1989) e Hsing *et al.* (1988), surgem os mesmos tipos de leis limite (max-estáveis) para o máximo, convenientemente normalizado, mas surge um parâmetro adicional com interesse, o chamado índice extremal, fácil de relacionar com as excedências de níveis normalizados.

Definição 1.1 *A sucessão estacionária \mathbf{X} tem índice extremal θ , $\theta \in [0, 1]$, se para cada $\tau > 0$, existe uma sucessão de níveis normalizados, $\{u_n^{(\tau)}\}_{n \geq 1}$, tal que*

$$P(M_n \leq u_n^{(\tau)}) \rightarrow \exp(-\theta\tau),$$

com θ fixo e independente de τ .

O índice extremal da sucessão estacionária \mathbf{X} permite relacionar a distribuição limite do máximo, G , a partir da distribuição limite, \widehat{G} , do máximo de uma sucessão i.i.d. $\widehat{\mathbf{X}} = \{\widehat{X}_n\}_{n \geq 1}$ com f.d. marginal igual à de \mathbf{X} . Tem-se

$$\lim_{n \rightarrow +\infty} P(M_n \leq a_n x + b_n) = G(x) = \widehat{G}^\theta(x) = \left[\lim_{n \rightarrow +\infty} P(\widehat{M}_n \leq a_n x + b_n) \right]^\theta, \quad (1.5)$$

onde $\widehat{M}_n = \bigvee_{i=1}^n \widehat{X}_i$.

Os resultados apresentados podem ser reformulados para o mínimo das n primeiras variáveis de \mathbf{X} , dado que

$$\bigwedge_{i=1}^n X_i = - \bigvee_{i=1}^n (-X_i),$$

onde ' \bigwedge ' denota o mínimo.

1.2.2 Sucessões d -dimensionais

Muitos fenómenos reais que envolvem acontecimentos extremos são de natureza multivariada. Tal conduziu ao desenvolvimento da teoria de valores extremos multivariados com um percurso mais ou menos paralelo à do caso univariado, mas de forma mais lenta.

Na extensão da teoria de valores extremos univariada ao contexto multivariado surgem novas questões características do contexto multivariado, como por exemplo, a definição de observação extrema multivariada e a avaliação da dependência entre as v.a.'s de um vetor. Alguns dos resultados que expusemos na secção anterior foram naturalmente estendidos para sucessões de vetores $\{\mathbf{X}_n = (\mathbf{X}_{n1}, \dots, \mathbf{X}_{nd})\}_{n \geq 1}$, considerando o máximo e as operações entre vetores componente a componente, isto é,

$$\mathbf{M}_n = \max \{\mathbf{X}_1, \dots, \mathbf{X}_n\} = (M_{n1}, \dots, M_{nd})$$

com

$$M_{nj} = \bigvee_{i=1}^n X_{ij}, \quad j \in \{1, \dots, d\},$$

e dados $\mathbf{a} = (a_1, \dots, a_d)$, $\mathbf{b} = (b_1, \dots, b_d)$, $\mathbf{x} = (x_1, \dots, x_d)$ em \mathbb{R}^d e $c \in \mathbb{R}$, tem-se $\mathbf{ax} + \mathbf{cb} = (a_1x_1 + cb_1, \dots, a_dx_d + cb_d)$, $\mathbf{a}^c = (a_1^c, \dots, a_d^c)$ e $\mathbf{a} \leq \mathbf{b}$ se e só se $a_j \leq b_j$, para todo $j \in \{1, \dots, d\}$.

De modo análogo ao caso univariado, se $\{\mathbf{X}_n\}_{n \geq 1}$ for uma sucessão d -dimensional de vetores aleatórios i.i.d. com f.d. F e se existirem sucessões $\{\mathbf{a}_n = (a_{n1}, \dots, a_{nd})\}_{n \geq 1}$ com $a_{nj} > 0$, $j \in \{1, \dots, d\}$, $n \geq 1$, e $\{\mathbf{b}_n = (b_{n1}, \dots, b_{nd})\}_{n \geq 1}$, tais que

$$P(\mathbf{M}_n \leq \mathbf{a}_n \mathbf{x} + \mathbf{b}_n) \xrightarrow[n \rightarrow +\infty]{w} G(\mathbf{x}), \quad (1.6)$$

com G não degenerada, então G diz-se uma distribuição multivariada de valores extremos (MEV). Quando tal acontece escrevemos $F \in D(G)$, isto é, F pertence ao domínio de atração (multivariado) de G (para o máximo).

A identificação das possíveis leis limite em (1.6) é dada pelo Teorema dos Tipos Extremos Multivariados, que se deve a Deheuvels (1978).

Sendo $G_j, j \in \{1, \dots, d\}$, as distribuições marginais de G , tem-se

$$P(M_{nj} \leq a_{nj}x_j + b_{nj}) \xrightarrow[n \rightarrow +\infty]{w} G_j(x_j), \quad (1.7)$$

onde G_j é do tipo extremo.

A classe de leis limite dada em (1.6) coincide com a classe das distribuições max-estáveis, isto é, existem $\mathbf{A}_k = (A_{k1}, \dots, A_{kd}) > \mathbf{0}$ e $\mathbf{B}_k = (B_{k1}, \dots, B_{kd}), k \geq 1$, tais que

$$G^k(\mathbf{A}_k \mathbf{x} + \mathbf{B}_k) = G(\mathbf{x}), \quad k \geq 1.$$

Contrariamente ao que sucede no caso univariado, no contexto multivariado não existe uma representação paramétrica geral para as distribuições multivariadas de valores extremos, devido à vasta classe de estruturas de dependência que podem ocorrer entre as margens do vetor (Y_1, \dots, Y_d) com distribuição G , o que veio dificultar a inferência estatística.

Supondo a existência de uma função de dependência (ou função cópula) e com base em restrições nesta função, várias sub-famílias paramétricas foram propostas na literatura (Joe (1997), Nelson (2006)).

Chamamos função de dependência de G a uma função de distribuição d -dimensional, D_G , com suporte $[0, 1]^d$, crescente em cada uma das suas variáveis e tal que

$$D_G(G_1(x_1), \dots, G_d(x_d)) = G(x_1, \dots, x_d), \quad (x_1, \dots, x_d) \in \mathbb{R}^d. \quad (1.8)$$

As margens de uma distribuição MEV G são contínuas e portanto D_G é uma função de distribuição com margens uniformemente distribuídas em $[0, 1]$. Nesta situação, tem-se a continuidade de G e de D_G .

Considerando a transformação $G_i(x_i) = u_i, i \in \{1, \dots, d\}$, podemos reescrever (1.8) da seguinte forma

$$D_G(u_1, \dots, u_d) = G(G_1^{-1}(u_1), \dots, G_d^{-1}(u_d)), \quad (u_1, \dots, u_d) \in [0, 1]^d.$$

Uma distribuição MEV, G , caracteriza-se por ter margens do tipo extremo e

$$D_G^k(\mathbf{u}^{1/k}) = D_G(\mathbf{u}), \quad k \geq 1, \quad \mathbf{u} \in [0, 1]^d. \quad (1.9)$$

Atendendo a que cada um dos três tipos da distribuição GEV pode ser transformado noutro, não há perda de generalidade em assumirmos que Y_1, \dots, Y_d possuem a mesma distribuição. Assumindo que para cada $j \in \{1, \dots, d\}$, G_j é Fréchet unitária, $G_j(u) = \exp(-1/u), u > 0$, é possível caracterizar a MEV G a partir da função

expoente. De facto, considerando

$$U_j = \left\{ 1 + \frac{\gamma_j(Y_j - \lambda_j)}{\delta_j} \right\}^{\frac{1}{\gamma_j}}, \quad j \in \{1, \dots, d\},$$

onde γ_j representa o índice de cauda e $\lambda_j \in \mathbb{R}$, $\delta_j > 0$ os parâmetros de localização e escala, respetivamente, tem-se que, para cada $j \in \{1, \dots, d\}$, a v.a. U_j possui distribuição Fréchet unitária e

$$P(U_1 \leq u_1, \dots, U_d \leq u_d) = G(u_1, \dots, u_d) = \exp(-V(u_1, \dots, u_d)), \quad u_j \in \mathbb{R}, \quad (1.10)$$

onde V é a função expoente definida do seguinte modo

$$V(u_1, \dots, u_d) = \int_{S_d} \bigvee_{i=1}^d \left(\frac{w_i}{u_i} \right) dH(w_1, \dots, w_d),$$

com H uma medida finita sobre $S_d = \{\mathbf{y} \in [0, +\infty)^d : y_1 + \dots + y_d = 1\}$ tal que $\int_{S_d} w_j dH(w_1, \dots, w_d) = 1$, para cada $j \in \{1, \dots, d\}$ (Pickands (1981)).

A função expoente satisfaz, entre outras, as seguintes propriedades

- (i) $V(\alpha u_1, \dots, \alpha u_d) = \alpha^{-1} V(u_1, \dots, u_d)$, $\alpha > 0$, $u_j \in \mathbb{R}^+$, $j \in \{1, \dots, d\}$;
- (ii) $V(+\infty, \dots, u_j, \dots, +\infty) = u_j^{-1}$, $u_j \in \mathbb{R}^+$, $j \in \{1, \dots, d\}$.

Para qualquer f.d. MEV G com f.d.'s marginais univariadas G_j , $j \in \{1, \dots, d\}$, tem-se

$$\prod_{j=1}^d G_j(x_j) \leq G(x_1, \dots, x_d) \leq \min_{j=1, \dots, d} \{G_j(x_j)\},$$

para todo $x_j \in \mathbb{R}$. O limite inferior corresponde ao caso em que as margens da f.d. MEV são independentes e o superior quando se tem a total dependência.

Em situações onde não se tem nenhum destes casos limite, é necessário ter coeficientes para quantificar o grau de dependência entre as margens da f.d. MEV.

A quantificação da dependência entre as margens de uma MEV é um tema que tem suscitado muito interesse na teoria de valores extremos. Vários coeficientes têm surgido na literatura para medir esta dependência.

Tiago de Oliveira (1962/63) introduziu o coeficiente de dependência extremal ε para distribuições bivariadas de valores extremos (BEV). Admitindo que o vetor (Y_i, Y_j) , $i, j \in \mathbb{N}$, tem distribuição BEV G de margens iguais a F , o coeficiente de dependência extremal é o real $\varepsilon(\{i, j\}) \in [1, 2]$ satisfazendo

$$P(Y_i \leq x, Y_j \leq x) = G(x, x) = F(x)^{\varepsilon(\{i, j\})}, \quad x \in \mathbb{R}. \quad (1.11)$$

Considerando que o vetor $(Y_i, Y_j), i, j \in \mathbb{N}$, tem uma qualquer f.d. G , não necessariamente BEV, mas com margens iguais a F , Buishand (1984) obteve em (1.11) a função $\varepsilon_{\{i,j\}}(x), x \in \mathbb{R}$, a qual se reduz ao coeficiente extremal $\varepsilon(\{i, j\})$ quando G é BEV.

Smith (1990) generalizou o coeficiente (1.11) às distribuições MEV como a constante $\varepsilon(\{1, \dots, d\}) \in [1, d]$ tal que

$$G(x, \dots, x) = F^{\varepsilon(\{1, \dots, d\})}(x), \quad x \in \mathbb{R}. \quad (1.12)$$

Este coeficiente relaciona-se com a função expoente do seguinte modo

$$\varepsilon(\{1, \dots, d\}) = V(\mathbf{1}),$$

onde $\mathbf{1}$ é o vetor unitário d -dimensional e permite quantificar a dependência extremal entre as margens univariadas da MEV G . Tem-se $\varepsilon(\{1, \dots, d\}) = 1$ se e só se G tem margens univariadas totalmente dependentes e $\varepsilon(\{1, \dots, d\}) = d$ se e só as margens univariadas de G são independentes.

A avaliação da dependência entre margens multivariadas de uma MEV é abordada em Ferreira (2011) através da introdução de novos coeficientes. O coeficiente $\varepsilon_1(\mathbf{A}, \mathbf{B})$ com $\mathbf{A} = \{1, 2, \dots, p\}$ e $\mathbf{B} = \{p+1, p+2, \dots, d\}$, permite avaliar a contribuição do subvetor $Y_{\mathbf{A}} = (Y_1, \dots, Y_p)$ com $p \leq d$, na estrutura de dependência do vetor $\mathbf{Y} = (Y_1, \dots, Y_d) \sim G$, sendo definido por

$$\varepsilon_1(\mathbf{A}, \mathbf{B}) = \frac{\varepsilon(\mathbf{A} \cup \mathbf{B})}{\varepsilon(\mathbf{A})} = \frac{\varepsilon(\{1, \dots, d\})}{\varepsilon(\{1, \dots, p\})}, \quad (1.13)$$

o qual varia entre 1 e $d - p + 1$. No caso em que $|\mathbf{A}| = |\mathbf{B}| = 1$, onde $|\mathbf{J}|$ representa o cardinal do conjunto \mathbf{J} , obtemos o coeficiente extremal ε de Tiago de Oliveira (1962/63) para distribuições BEV.

Com vista à quantificação da dependência entre os vetores $Y_{\mathbf{A}}$ e $Y_{\mathbf{B}}$, com $Y_{\mathbf{B}} = (Y_{p+1}, \dots, Y_d)$, Ferreira (2011) propôs o coeficiente $\varepsilon_2(\mathbf{A}, \mathbf{B})$ definido por,

$$\varepsilon_2(\mathbf{A}, \mathbf{B}) = \frac{\varepsilon(\{1, \dots, d\})}{\varepsilon(\{1, \dots, p\}) + \varepsilon(\{p+1, \dots, d\})}. \quad (1.14)$$

Quando $|\mathbf{A}| = |\mathbf{B}|$ e $G_{Y_{\mathbf{A}}} = G_{Y_{\mathbf{B}}}$, tem-se

$$\varepsilon_2(\mathbf{A}, \mathbf{B}) = \frac{1}{2} \varepsilon_1(\mathbf{A}, \mathbf{B})$$

e, no caso particular de $|\mathbf{A}| = |\mathbf{B}| = 1$, obtém-se

$$\varepsilon_2(\mathbf{A}, \mathbf{B}) = \frac{1}{2}\varepsilon(\mathbf{A} \cup \mathbf{B}),$$

que a autora designa por coeficiente extremal redimensionado. Este coeficiente toma valores em $[1/2, 1]$. Tem-se $\varepsilon_2(\mathbf{A}, \mathbf{B}) = 1$ se e só se $Y_{\mathbf{A}}$ e $Y_{\mathbf{B}}$ são independentes. No caso de total dependência entre $Y_{\mathbf{A}}$ e $Y_{\mathbf{B}}$ o coeficiente assume o valor $1/2$.

Uma forma de avaliar a dependência entre $Y_{\mathbf{A}}$ e $Y_{\mathbf{B}}$ em hiper-octantes determinados por caudas superiores é dada em Li (2009) (ver também Ferreira (2011)) através da introdução do coeficiente $\lambda(\mathbf{A}, \mathbf{B})$ definido por

$$\lambda(\mathbf{A}, \mathbf{B}) = \lim_{x \rightarrow x_F} P(Y_{\mathbf{A}} > x_{\mathbf{A}} | Y_{\mathbf{B}} > x_{\mathbf{B}}), \quad (1.15)$$

onde $\{Y_{\mathbf{C}} > x_{\mathbf{C}}\}$ denota o acontecimento $\bigcap_{i \in \mathbf{C}} \{Y_i > x\}$, desde que o limite exista.

O coeficiente $\lambda(\mathbf{A}, \mathbf{B})$ dá-nos uma medida, com valores em $[0, 1]$, para a dependência entre $Y_{\mathbf{A}}$ e $Y_{\mathbf{B}}$ sobre $\{\mathbf{x} \in S_G : x \text{ elevado}\}$, onde S_G denota o suporte da distribuição G e \mathbf{x} é o vetor d -dimensional com todas as componentes iguais a x . Este coeficiente toma os valores 0 ou 1 quando na cauda $Y_{\mathbf{A}}$ e $Y_{\mathbf{B}}$ são independentes ou totalmente dependentes, respetivamente.

Se para alguns subvetores $Y_{\mathbf{A}}$ e $Y_{\mathbf{B}}$, com $\mathbf{A} \cup \mathbf{B} = \{1, \dots, d\}$ e $\mathbf{A} \cap \mathbf{B} = \emptyset$, este coeficiente é positivo, então o vetor \mathbf{Y} diz-se dependente na cauda multivariada. Quando o coeficiente é nulo para quaisquer subvetores, o vetor \mathbf{Y} diz-se independente na cauda (Schmidt (2002)).

Quando em (1.15) consideramos $|\mathbf{A}| = |\mathbf{B}| = 1$ obtemos o coeficiente de dependência na cauda superior de G para o caso bivariado (Sibuya (1960)).

No contexto bivariado o coeficiente de dependência na cauda superior e o coeficiente de dependência extremal estão relacionados da seguinte forma

$$\lambda(\{i\}, \{j\}) = 2 - \varepsilon(\{i, j\}). \quad (1.16)$$

Qualquer um dos coeficientes referidos anteriormente não mede corretamente a dependência na distribuição limite do máximo quando a sucessão

$$\{\mathbf{X}_n = (\mathbf{X}_{n1}, \dots, \mathbf{X}_{nd})\}_{n \geq 1}$$

possui índice extremal multivariado (Nandagopalan (1990)) diferente de um. Em Martins e Ferreira (2005) e Viseu *et al.* (2012) são propostos ajustamentos destes coeficientes de modo a englobar esse caso.

1.2.3 Campos aleatórios

O estudo de propriedades probabilísticas associadas a modelos que descrevem fenómenos aleatórios que variam no tempo e no espaço, de que são exemplo fenómenos aleatórios hidrológicos ou climatológicos, esteve na base do desenvolvimento da teoria de valores extremos para campos aleatórios $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbf{I}}$, com $\mathbf{I} \subseteq \mathbb{R}^d$, $d \in \mathbb{N} \setminus \{1\}$.

O estudo do comportamento extremal de campos aleatórios discretos $\{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{N}^d}$, com origem principalmente em Adler (1981) e Piterbarg (1996), seguiu uma direção de certo modo paralela à do caso de sucessões univariadas, tendo apresentado um forte incremento na última década, cujo reflexo podemos ver, por exemplo, em Pereira e Ferreira (2006), Ferreira e Pereira (2008, 2012), Pereira (2009, 2010), entre outros.

Uma classe de campos aleatórios que tem suscitado grande interesse no estudo de valores extremos é a dos campos max-estáveis, uma vez que, teoricamente, se justificam os adequados para a modelação de valores extremos espaciais.

Definição 1.2 *Um campo aleatório $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbf{I}}$, $\mathbf{I} \subseteq \mathbb{R}^d$, é max-estável se para $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}$, $n \geq 1$, cópias i.i.d. de \mathbf{X} existirem funções reais $a_n(\mathbf{i}) > 0$ e $b_n(\mathbf{i})$, tais que*

$$\left\{ \bigvee_{j=1}^n \frac{X_{\mathbf{i}}^{(j)} - b_n(\mathbf{i})}{a_n(\mathbf{i})} \right\}_{\mathbf{i} \in \mathbf{I}} \stackrel{d}{=} \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbf{I}}$$

onde $\stackrel{d}{=}$ denota a igualdade em distribuição.

A caracterização do domínio de atração dos campos max-estáveis tem interessado muitos autores. A este respeito salientamos os trabalhos de de Haan e Lin (1991), de Haan e Ferreira (2006), entre outros.

Definição 1.3 *Diz-se que o campo $\mathbf{Y} = \{Y_{\mathbf{i}}\}_{\mathbf{i} \in \mathbf{I}}$, $\mathbf{I} \subseteq \mathbb{R}^d$, pertence ao domínio de atração de $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbf{I}}$ se, para $\mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(n)}$, $n \geq 1$, cópias i.i.d. de \mathbf{Y} e para as funções reais $a_n(\mathbf{i}) > 0$, $b_n(\mathbf{i})$,*

$$\left\{ \bigvee_{j=1}^n \frac{Y_{\mathbf{i}}^{(j)} - b_n(\mathbf{i})}{a_n(\mathbf{i})} \right\}_{\mathbf{i} \in \mathbf{I}} \xrightarrow[n \rightarrow +\infty]{d} \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbf{I}}, \quad (1.17)$$

onde $\xrightarrow[n \rightarrow +\infty]{d}$ denota a convergência em distribuição, admitindo que para cada $\mathbf{i} \in \mathbf{I}$, $X_{\mathbf{i}}$ é não degenerada.

Como consequência das definições anteriores tem-se que o campo limite em (1.17) é necessariamente um campo max-estável e que as distribuições marginais univariadas e multivariadas de um campo max-estável são, respetivamente, GEV e MEV.

A necessidade de modelar fenômenos naturais, nomeadamente meteorológicos e hidrológicos, conduziu ao desenvolvimento de métodos para definir e simular classes de campos max-estáveis. Existem na literatura várias classes de campos max-estáveis, sendo as mais referenciadas nas aplicações, a dos campos max-móveis (de Haan e Pickands (1986)) e a que se obtém como limite do máximo de certos campos Gaussianos (Brown e Resnick (1977), Kabluchko (2009), Smith (1990), de Haan e Ferreira (2006), Schlather e Tawn (2003)). Alguns destes modelos encontram-se implementados nas recentes *packages RandomFields* e *SpatialExtremes* integrados no programa *R* e desenvolvidas por Schlather (2012) e Ribatet (2009), respetivamente.

Nos campos aleatórios Gaussianos a função de correlação permite caracterizar a sua estrutura de dependência (Abrahamsen (1997)). Contudo, apesar da sua vasta aplicação, não se adaptam à modelação de valores extremos, surgindo assim a necessidade de encontrar formas de avaliar a dependência entre extremos espaciais.

O problema da quantificação da dependência entre extremos de campos max-estáveis tem sido objeto de estudo por parte de vários autores. A motivação quase sempre é a de se prevenir contra algum acontecimento extremo cujas consequências possam ser indesejáveis e desastrosas.

Começamos por destacar o trabalho de Schlather e Tawn (2003) onde se introduz a função coeficiente extremal para avaliar a dependência entre as margens de distribuições bivariadas de um campo aleatório estacionário e max-estável com f.d. marginal univariada de Fréchet unitária.

Admitindo que $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{R}^d}$ é um campo aleatório estacionário max-estável com f.d. marginal univariada F de Fréchet unitária, $F(x) = \exp(-\frac{1}{x})$, $x > 0$, a função coeficiente extremal, $\varepsilon(\mathbf{h}) : \mathbb{R}^d \rightarrow [1, 2]$, satisfaz

$$\begin{aligned} G_{\mathbf{i}, \mathbf{i}+\mathbf{h}}(x, x) &= P(X_{\mathbf{i}} \leq x, X_{\mathbf{i}+\mathbf{h}} \leq x) = F^{\varepsilon(\mathbf{h})}(x) = \\ &= \exp\left(-\frac{\varepsilon(\mathbf{h})}{x}\right), \mathbf{i} \in \mathbb{R}^d, x > 0. \end{aligned}$$

Tem-se $\varepsilon(\mathbf{h}) = 1$ se e só se $G_{\mathbf{i}, \mathbf{i}+\mathbf{h}}$ tem margens totalmente dependentes e $\varepsilon(\mathbf{h}) = 2$ se e só se $G_{\mathbf{i}, \mathbf{i}+\mathbf{h}}$ tem margens independentes.

Observação 1.1 *Se o campo aleatório max-estável \mathbf{X} não é estacionário, a avaliação da dependência entre as margens de uma distribuição bivariada de \mathbf{X} , $G_{\mathbf{i}, \mathbf{i}+\mathbf{h}}$, é dada pelo real $\varepsilon(\{\mathbf{i}, \mathbf{i} + \mathbf{h}\}) \in [1, 2]$ satisfazendo*

$$\begin{aligned} G_{\mathbf{i}, \mathbf{i}+\mathbf{h}}(x, x) &= P(X_{\mathbf{i}} \leq x, X_{\mathbf{i}+\mathbf{h}} \leq x) = F^{\varepsilon(\{\mathbf{i}, \mathbf{i}+\mathbf{h}\})}(x) = \\ &= \exp\left(-\frac{\varepsilon(\{\mathbf{i}, \mathbf{i} + \mathbf{h}\})}{x}\right), \mathbf{i}, \mathbf{h} \in \mathbb{R}^d, x > 0. \end{aligned}$$

Se o campo aleatório max-estável e estacionário, \mathbf{X} , com f.d. marginal univari-

ada, F , de Fréchet unitária for ainda isotrópico, então

$$G_{\mathbf{i}, \mathbf{i}+\mathbf{h}}(u_1, u_2) = \exp(-V_h(u_1, u_2)),$$

onde $h = \|\mathbf{h}\|$ e

$$V_h(u_1, u_2) = 2 \int_0^1 \max\left(\frac{w}{u_1}, \frac{1-w}{u_2}\right) dH_h(w) \quad (1.18)$$

com H_h uma f.d. sobre $[0, 1]$ tal que $\int_0^1 w dH_h(w) = 0.5$. Nestas condições, a função coeficiente extremal vem dependente da distância h , isto é,

$$G_{\mathbf{i}, \mathbf{i}+\mathbf{h}}(x, x) = P(X_{\mathbf{i}} \leq x, X_{\mathbf{i}+\mathbf{h}} \leq x) = P^{\varepsilon(h)}(X_{\mathbf{i}} \leq x), \quad x > 0,$$

sendo $\varepsilon(h) = V_h(1, 1)$. Quanto maior for a distância h entre $\mathbf{i}, \mathbf{i} + \mathbf{h}$, menor tenderá a ser a dependência entre as margens de $G_{\mathbf{i}, \mathbf{i}+\mathbf{h}}$.

Uma outra abordagem à dependência extremal em campos aleatórios max-estáveis, estacionários e isotrópicos teve origem no conceito de semivariograma definido por

$$\gamma(h) = \frac{1}{2} E[(X_{\mathbf{i}} - X_{\mathbf{i}+\mathbf{h}})^2], \quad \mathbf{i}, \mathbf{h} \in \mathbb{R}^d, \quad h = \|\mathbf{h}\|,$$

amplamente utilizado em Geoestatística.

A sua inadequação no estudo da dependência entre extremos de campos max-estáveis, uma vez que se pode ter $E(X_{\mathbf{j}}^2) = +\infty$, motivou a introdução do madograma normalizado (Cooley (2005)), definido por

$$\nu_F(h) = \frac{1}{2} E[|F(X_{\mathbf{i}}) - F(X_{\mathbf{i}+\mathbf{h}})|], \quad \mathbf{i}, \mathbf{h} \in \mathbb{R}^d, \quad h = \|\mathbf{h}\|, \quad (1.19)$$

relacionando-se com $\varepsilon(h)$, da seguinte forma

$$\nu_F(h) = \frac{\varepsilon(h)}{1 + \varepsilon(h)} - \frac{1}{2}.$$

Os coeficientes $\varepsilon(h)$ e $\nu_F(h)$ são indicados na avaliação da dependência entre os acontecimentos $\{X_{\mathbf{i}} \leq u\}$ e $\{X_{\mathbf{i}+\mathbf{h}} \leq v\}$ apenas no caso em que $u = v$. Para contornar esta limitação Naveau *et al.* (2009) introduzem o λ -madograma definido por

$$\nu(\lambda, h) = \frac{1}{2} E[|F^\lambda(X_{\mathbf{i}}) - F^{1-\lambda}(X_{\mathbf{i}+\mathbf{h}})|], \quad 0 < \lambda < 1, \quad \mathbf{i}, \mathbf{h} \in \mathbb{R}^d, \quad h = \|\mathbf{h}\|. \quad (1.20)$$

Este coeficiente, com valores no intervalo $[0, \frac{1}{2}]$ relaciona-se com a função expo-

ente V_h (1.18) da seguinte forma

$$\nu(\lambda, h) = \frac{V_h(\lambda, 1 - \lambda)}{1 + V_h(\lambda, 1 - \lambda)} - \frac{3}{2(1 + \lambda(2 - \lambda))}.$$

Tal como já referimos anteriormente, a *package SpatialExtremes* do programa *R* tem disponível alguns modelos max-estáveis e permite estimar alguns dos coeficientes aqui apresentados, nomeadamente o λ -madograma (Ribatet (2009)).

Capítulo 2

Matrizes de coeficientes de dependência

2.1 Introdução

Nos problemas espaciais e multivariados de extremos é importante analisar a estrutura de dependência para se perceber, por exemplo, em que medida a ocorrência de valores extremos em determinadas localizações influencia o comportamento da variável em localizações vizinhas. Em termos práticos, nem sempre é possível aplicar modelos com estrutura de dependência conhecida sendo, por isso, relevante estudar medidas de dependência que a permitam avaliar. No primeiro capítulo foram elencadas algumas das medidas de dependência existentes na literatura e que têm sido amplamente utilizadas em aplicações práticas.

Neste capítulo vamos considerar um campo aleatório discreto $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{Z}^2}$ max-estável, com f.d. marginal univariada F . Se o campo \mathbf{X} é isotrópico as medidas de dependência extremal $\varepsilon(\{\mathbf{i}, \mathbf{j}\})$ e $\lambda(\{\mathbf{i}\}, \{\mathbf{j}\})$ apenas dependem da distância, $\|\mathbf{i} - \mathbf{j}\|$, entre as localizações \mathbf{i} e \mathbf{j} . Contudo, em grande parte das situações reais não temos isotropia, o que exige a avaliação da dependência espacial em diferentes direções de \mathbb{Z}^2 . Foi esta questão que motivou a introdução das matrizes de coeficientes de dependência multivariada na cauda superior e extremas (Fonseca *et al.* (2009)), onde cada elemento fornece uma medida de dependência entre as variáveis indexadas num conjunto \mathbf{A} e as variáveis indexadas num conjunto de localizações vizinhas de \mathbf{A} , segundo uma dada direção.

Nesta abordagem multidirecional generaliza-se (1.16) e obtêm-se algumas propriedades que revelam relações importantes entre os coeficientes de dependência multivariada. Estas matrizes serão utilizadas para estudar a dependência em três campos aleatórios max-estáveis.

O cálculo do coeficiente de dependência na cauda superior $\lambda(\mathbf{B}, \mathbf{A})$ definido em

(1.15), com \mathbf{A} e \mathbf{B} conjuntos de localizações de grande dimensão, pode ser uma tarefa árdua, pelo que propomos um algoritmo para a sua determinação e aplicamo-lo em dois exemplos.

Finalizamos este capítulo com a introdução de um campo max-estável e anisotrópico de máximos móveis de máximos multivariados (M4) e que utilizaremos nos estudos de simulação nos Capítulos 3 e 4. Para este campo, será avaliada a dependência na cauda superior entre as variáveis indexadas num conjunto \mathbf{A} e as variáveis indexadas num conjunto de localizações vizinhas de \mathbf{A} .

2.2 Dependência local em eventos extremos

No sentido de simplificar a leitura e a escrita, passaremos a denotar $\bigcap_{i \in \mathbf{C}} \{X_i > x\}$ e $\bigcap_{i \in \mathbf{C}} \{X_i \leq x\}$ por $\{X_{\mathbf{C}} > x_{\mathbf{C}}\}$ e $\{X_{\mathbf{C}} \leq x_{\mathbf{C}}\}$, $\mathbf{C} \subseteq \mathbb{Z}^2$, $x \in \mathbb{R}$, respetivamente. Apresentam-se, de seguida, algumas relações entre o coeficiente $\lambda(\mathbf{B}, \mathbf{A})$ e o coeficiente extremal de dependência multivariada ε .

Se \mathbf{X} é um campo aleatório max-estável então, quaisquer que sejam os conjuntos de localizações \mathbf{A} e \mathbf{B} , que também designaremos por regiões \mathbf{A} e \mathbf{B} , tem-se

$$\begin{aligned}
 \lambda(\mathbf{B}, \mathbf{A}) &= \lim_{x \rightarrow x^F} \frac{P(X_{\mathbf{B}} > x_{\mathbf{B}}, X_{\mathbf{A}} > x_{\mathbf{A}})}{P(X_{\mathbf{A}} > x_{\mathbf{A}})} \\
 &= \lim_{x \rightarrow x^F} \frac{1 - \sum_{\phi \neq \mathbf{J} \subseteq \mathbf{A} \cup \mathbf{B}} (-1)^{|\mathbf{J}|+1} P(X_{\mathbf{J}} \leq x_{\mathbf{J}})}{1 - \sum_{\phi \neq \mathbf{J} \subseteq \mathbf{A}} (-1)^{|\mathbf{J}|+1} P(X_{\mathbf{J}} \leq x_{\mathbf{J}})} \\
 &= \frac{\sum_{\phi \neq \mathbf{J} \subseteq \mathbf{A} \cup \mathbf{B}} (-1)^{|\mathbf{J}|+1} \varepsilon(\mathbf{J})}{\sum_{\phi \neq \mathbf{J} \subseteq \mathbf{A}} (-1)^{|\mathbf{J}|+1} \varepsilon(\mathbf{J})}. \tag{2.1}
 \end{aligned}$$

Deste modo, o coeficiente de dependência na cauda reflete a dependência entre X_i , $i \in \mathbf{J}$, para cada $\mathbf{J} \subseteq \mathbf{A} \cup \mathbf{B}$. No caso particular em que $\mathbf{A} = \{\mathbf{i}\}$ e $\mathbf{B} = \{\mathbf{j}\}$ obtém-se (1.16).

Para evidenciar a contribuição de $\varepsilon(\mathbf{A})$, $\varepsilon(\mathbf{B})$ e $\varepsilon(\mathbf{A} \cup \mathbf{B})$ neste coeficiente reescrevemos (2.1) da seguinte forma:

$$\lambda(\mathbf{B}, \mathbf{A}) = \frac{\sum_{\phi \neq \mathbf{J} \subseteq \mathbf{A} \cup \mathbf{B}} (-1)^{|\mathbf{J}|+1} \frac{\varepsilon(\mathbf{J})}{\varepsilon(\mathbf{A} \cup \mathbf{B})}}{\sum_{\phi \neq \mathbf{J} \subseteq \mathbf{A}} (-1)^{|\mathbf{J}|+1} \frac{\varepsilon(\mathbf{J})}{\varepsilon(\mathbf{A})}} \times \varepsilon_1(\mathbf{A}, \mathbf{B}) \tag{2.2}$$

2.2. DEPENDÊNCIA LOCAL EM EVENTOS EXTREMOS

$$= \frac{\sum_{\phi \neq \mathbf{J} \subseteq \mathbf{A} \cup \mathbf{B}} (-1)^{|\mathbf{J}|+1} \varepsilon(\mathbf{J}) \frac{\varepsilon(\mathbf{A}) + \varepsilon(\mathbf{B})}{\varepsilon(\mathbf{A} \cup \mathbf{B})}}{\sum_{\phi \neq \mathbf{J} \subseteq \mathbf{A}} (-1)^{|\mathbf{J}|+1} \varepsilon(\mathbf{J})} \times \varepsilon_2(\mathbf{A}, \mathbf{B}), \quad (2.3)$$

onde $\varepsilon_1(\mathbf{A}, \mathbf{B})$ e $\varepsilon_2(\mathbf{A}, \mathbf{B})$ são os coeficientes de dependência (1.13) e (1.14) introduzidos em Ferreira (2011). Estes coeficientes avaliam o grau de dependência entre os eventos $\{X_{\mathbf{A}} \leq x_{\mathbf{A}}\}$ e $\{X_{\mathbf{B}} \leq x_{\mathbf{B}}\}$, tendo sido motivados pelas seguintes relações

$$P(X_{\mathbf{A} \cup \mathbf{B}} \leq x_{\mathbf{A} \cup \mathbf{B}}) = P^{\frac{\varepsilon(\mathbf{A} \cup \mathbf{B})}{\varepsilon(\mathbf{A})}}(X_{\mathbf{A}} \leq x_{\mathbf{A}})$$

e

$$P(X_{\mathbf{A} \cup \mathbf{B}} \leq x_{\mathbf{A} \cup \mathbf{B}}) = (P(X_{\mathbf{A}} \leq x_{\mathbf{A}}) P(X_{\mathbf{B}} \leq x_{\mathbf{B}}))^{\frac{\varepsilon(\mathbf{A} \cup \mathbf{B})}{\varepsilon(\mathbf{A}) + \varepsilon(\mathbf{B})}}.$$

Uma interpretação mais clara das relações (2.2) e (2.3) surgirá na proposição que se segue, onde será considerada a seguinte notação

$$\alpha(\mathbf{C}) = \lim_{x \rightarrow x^F} \frac{P(X_{\mathbf{C}} > x_{\mathbf{C}})}{1 - P(X_{\mathbf{C}} \leq x_{\mathbf{C}})}$$

e

$$\beta(\mathbf{C}, \mathbf{D}) = \lim_{x \rightarrow x^F} \frac{P(X_{\mathbf{C}} > x_{\mathbf{C}} \vee X_{\mathbf{D}} > x_{\mathbf{D}})}{1 - P(X_{\mathbf{C}} \leq x_{\mathbf{C}}, X_{\mathbf{D}} \leq x_{\mathbf{D}})}.$$

Refira-se que $\alpha(\mathbf{A})$ e $\alpha(\mathbf{A} \cup \mathbf{B})$ podem ser relacionados com $\lambda(\mathbf{B}, \mathbf{A})$, através de (2.2) e (2.3), uma vez que para qualquer conjunto de localizações \mathbf{C} ,

$$\alpha(\mathbf{C}) = \sum_{\phi \neq \mathbf{J} \subseteq \mathbf{C}} (-1)^{|\mathbf{J}|+1} \frac{\varepsilon(\mathbf{J})}{\varepsilon(\mathbf{C})}.$$

Proposição 2.1 *Seja \mathbf{X} um campo aleatório max-estável sobre \mathbb{Z}^2 e \mathbf{A} e \mathbf{B} dois conjuntos de localizações em \mathbb{Z}^2 . Então,*

1.

$$\lambda(\mathbf{B}, \mathbf{A}) = 1 + \frac{\alpha(\mathbf{B}) \varepsilon(\mathbf{B})}{\alpha(\mathbf{A}) \varepsilon(\mathbf{A})} - \frac{\beta(\mathbf{A}, \mathbf{B})}{\alpha(\mathbf{A})} \varepsilon_1(\mathbf{A}, \mathbf{B}); \quad (2.4)$$

2.

$$\lambda(\mathbf{B}, \mathbf{A}) = \frac{\alpha(\mathbf{A} \cup \mathbf{B})}{\alpha(\mathbf{A})} \varepsilon_1(\mathbf{A}, \mathbf{B}). \quad (2.5)$$

Demonstração:

1. Atendendo à definição de coeficiente de dependência na cauda, obtém-se

$$\lambda(\mathbf{B}, \mathbf{A}) = \lim_{x \rightarrow x^F} \frac{P(X_{\mathbf{B}} > x_{\mathbf{B}}) + P(X_{\mathbf{A}} > x_{\mathbf{A}}) - P(X_{\mathbf{B}} > x_{\mathbf{B}} \vee X_{\mathbf{A}} > x_{\mathbf{A}})}{P(X_{\mathbf{A}} > x_{\mathbf{A}})}$$

$$\begin{aligned}
 &= \lim_{x \rightarrow x^F} 1 + \frac{P(X_{\mathbf{B}} > x_{\mathbf{B}}) - P(X_{\mathbf{B}} > x_{\mathbf{B}} \vee X_{\mathbf{A}} > x_{\mathbf{A}})}{P(X_{\mathbf{A}} > x_{\mathbf{A}})} \\
 &= 1 + \frac{\alpha(\mathbf{B})}{\alpha(\mathbf{A})} \lim_{x \rightarrow x^F} \frac{1 - P(X_{\mathbf{B}} \leq x_{\mathbf{B}})}{1 - P(X_{\mathbf{A}} \leq x_{\mathbf{A}})} - \frac{\beta(\mathbf{A}, \mathbf{B})}{\alpha(\mathbf{A})} \lim_{x \rightarrow x^F} \frac{1 - P(X_{\mathbf{A} \cup \mathbf{B}} \leq x_{\mathbf{A} \cup \mathbf{B}})}{1 - P(X_{\mathbf{A}} \leq x_{\mathbf{A}})} \\
 &= 1 + \frac{\alpha(\mathbf{B}) \varepsilon(\mathbf{B})}{\alpha(\mathbf{A}) \varepsilon(\mathbf{A})} - \frac{\beta(\mathbf{A}, \mathbf{B})}{\alpha(\mathbf{A})} \varepsilon_1(\mathbf{A}, \mathbf{B}).
 \end{aligned}$$

2. Como

$$\beta(\mathbf{A}, \mathbf{B}) = \alpha(\mathbf{A}) \frac{\varepsilon(\mathbf{A})}{\varepsilon(\mathbf{A} \cup \mathbf{B})} + \alpha(\mathbf{B}) \frac{\varepsilon(\mathbf{B})}{\varepsilon(\mathbf{A} \cup \mathbf{B})} - \alpha(\mathbf{A} \cup \mathbf{B}),$$

o resultado segue-se de (2.4). \square

Note-se que $\alpha(\mathbf{A})$ e $\beta(\mathbf{A}, \mathbf{B})$, em geral, não são iguais a um, como é ilustrado nos exemplos apresentados em Ferreira (2011).

A igualdade (2.4) é a extensão ao contexto multivariado de (1.16). Por outro lado, a igualdade (2.5) estabelece uma nova relação entre $\lambda(\{\mathbf{i}\}, \{\mathbf{j}\})$ e $\varepsilon(\{\mathbf{i}, \mathbf{j}\})$, sendo

$$\lambda(\{\mathbf{i}\}, \{\mathbf{j}\}) = \alpha(\{\mathbf{i}, \mathbf{j}\}) \varepsilon(\{\mathbf{i}, \mathbf{j}\}),$$

com $\lambda(\{\mathbf{i}\}, \{\mathbf{j}\}) = \varepsilon(\{\mathbf{i}, \mathbf{j}\}) = 1$, isto é, $X_{\mathbf{i}}$ e $X_{\mathbf{j}}$ totalmente dependentes, se $\alpha(\{\mathbf{i}, \mathbf{j}\}) = 1$. Genericamente, se

$$P\left(\bigcap_{\mathbf{i} \in \mathbf{C}} \{X_{\mathbf{i}} > x\}\right) \sim P\left(\bigcup_{\mathbf{i} \in \mathbf{C}} \{X_{\mathbf{i}} > x\}\right), \text{ quando } x \rightarrow x_F,$$

para os conjuntos de localizações $\mathbf{C} = \mathbf{A}$ e $\mathbf{C} = \mathbf{A} \cup \mathbf{B}$, então tem-se

$$\lambda(\mathbf{B}, \mathbf{A}) = \varepsilon_1(\mathbf{A}, \mathbf{B}) = 1.$$

Seguidamente iremos descrever a dependência espacial através de matrizes de coeficientes de dependência, depois de introduzirmos os conceitos necessários.

Para cada $\mathbf{i} = (i_1, i_2) \in \mathbb{Z}^2$ sejam $s_j(\mathbf{i})$, $j \in \{1, 2, \dots, 8\}$, os vizinhos de \mathbf{i} definidos por

$$\begin{aligned}
 s_1(\mathbf{i}) &= (i_1 + 1, i_2), & s_2(\mathbf{i}) &= \mathbf{i} + \mathbf{1}, & s_3(\mathbf{i}) &= (i_1, i_2 + 1), & s_4(\mathbf{i}) &= (i_1 - 1, i_2 + 1), \\
 s_5(\mathbf{i}) &= (i_1 - 1, i_2), & s_6(\mathbf{i}) &= \mathbf{i} - \mathbf{1}, & s_7(\mathbf{i}) &= (i_1, i_2 - 1), & s_8(\mathbf{i}) &= (i_1 + 1, i_2 - 1),
 \end{aligned} \tag{2.6}$$

e consideremos $s_j^k(\mathbf{i}) = (s_j \circ \dots \circ s_j)(\mathbf{i})$, a função composta de s_j , k vezes, com $k \geq 1$ e $s_j^0(\mathbf{i}) = \mathbf{i}$, $j \in \{1, 2, \dots, 8\}$.

Definimos a fronteira do conjunto \mathbf{A} por

$$\mathfrak{B}(\mathbf{A}) = \{\mathbf{i} \in \mathbf{A} : s_j(\mathbf{i}) \notin \mathbf{A} \text{ para algum } j \in \{1, 2, \dots, 8\}\}$$

e seja

$$T_{s_j}^k(\mathbf{A}) = \{s_j^k(\mathbf{i}) : \mathbf{i} \in \mathfrak{B}(\mathbf{A})\} - \mathbf{A}, \quad j \in \{1, 2, \dots, 8\} \quad k \geq 1, \quad (2.7)$$

o conjunto formado pelas translações de passo $k \geq 1$, dos elementos de $\mathfrak{B}(\mathbf{A})$ segundo a direção s_j , ou seja, é o conjunto das localizações vizinhas de \mathbf{A} a passo k .

Considere-se a matriz $\Lambda(T_s^k(\mathbf{A}), \mathbf{A})$, de coeficientes de dependência multivariada na cauda superior, cujos elementos medem a tendência para a ocorrência simultânea de valores elevados de \mathbf{X} nas regiões \mathbf{A} e $T_s^k(\mathbf{A})$,

$$\Lambda(T_s^k(\mathbf{A}), \mathbf{A}) = \begin{bmatrix} \lambda(T_{s_4}^k(\mathbf{A}), \mathbf{A}) & \lambda(T_{s_3}^k(\mathbf{A}), \mathbf{A}) & \lambda(T_{s_2}^k(\mathbf{A}), \mathbf{A}) \\ \lambda(T_{s_5}^k(\mathbf{A}), \mathbf{A}) & \lambda(T_{s_0}^k(\mathbf{A}), \mathbf{A}) & \lambda(T_{s_1}^k(\mathbf{A}), \mathbf{A}) \\ \lambda(T_{s_6}^k(\mathbf{A}), \mathbf{A}) & \lambda(T_{s_7}^k(\mathbf{A}), \mathbf{A}) & \lambda(T_{s_8}^k(\mathbf{A}), \mathbf{A}) \end{bmatrix} \quad (2.8)$$

onde $T_{s_0}^k(\mathbf{A}) = \mathbf{A}$ e $\lambda(T_{s_0}^k(\mathbf{A}), \mathbf{A}) = \lambda(\mathbf{A}, \mathbf{A}) = 1$.

Como consequência da Proposição 2.1, a matriz $\Lambda(T_s^k(\mathbf{A}), \mathbf{A})$, $k \geq 1$, obtém-se a partir da matriz $\Lambda(T_s^{k-1}(\mathbf{A}), \mathbf{A})$, como se mostra no próximo resultado.

Corolário 2.1 *Seja \mathbf{X} um campo aleatório max-estável sobre \mathbb{Z}^2 e \mathbf{A} um conjunto de localizações em \mathbb{Z}^2 . Então, para quaisquer $t, j \in \{1, 2, \dots, 8\}$ e $k > 1$, tem-se*

1.
$$\begin{aligned} \lambda(T_{s_j}^k(\mathbf{A}), \mathbf{A}) &= \lambda(T_{s_t}^{k-1}(\mathbf{A}), \mathbf{A}) + \frac{\alpha(T_{s_j}^k(\mathbf{A})) \varepsilon(T_{s_j}^k(\mathbf{A}))}{\alpha(\mathbf{A}) \varepsilon(\mathbf{A})} - \\ &- \frac{\beta(T_{s_j}^k(\mathbf{A}), \mathbf{A})}{\alpha(\mathbf{A})} \varepsilon_1(\mathbf{A}, T_{s_j}^k(\mathbf{A})) - \\ &- \frac{\alpha(T_{s_t}^{k-1}(\mathbf{A})) \varepsilon(T_{s_t}^{k-1}(\mathbf{A}))}{\alpha(\mathbf{A}) \varepsilon(\mathbf{A})} + \frac{\beta(T_{s_t}^{k-1}(\mathbf{A}), \mathbf{A})}{\alpha(\mathbf{A})} \varepsilon_1(\mathbf{A}, T_{s_t}^{k-1}(\mathbf{A})); \end{aligned}$$
2.
$$\begin{aligned} \lambda(T_{s_j}^k(\mathbf{A}), \mathbf{A}) &= \lambda(T_{s_t}^{k-1}(\mathbf{A}), \mathbf{A}) + \frac{\alpha(\mathbf{A} \cup T_{s_j}^k(\mathbf{A}))}{\alpha(\mathbf{A})} \varepsilon_1(\mathbf{A}, T_{s_j}^k(\mathbf{A})) - \\ &- \frac{\alpha(\mathbf{A} \cup T_{s_t}^{k-1}(\mathbf{A}))}{\alpha(\mathbf{A})} \varepsilon_1(\mathbf{A}, T_{s_t}^{k-1}(\mathbf{A})); \end{aligned}$$
3.
$$\lambda(\{s_j^k(\mathbf{i})\}, \{\mathbf{i}\}) = \lambda(\{s_t^{k-1}(\mathbf{i})\}, \{\mathbf{i}\}) + \varepsilon(\{s_t^{k-1}(\mathbf{i}), \mathbf{i}\}) - \varepsilon(\{s_j^k(\mathbf{i}), \mathbf{i}\}), \quad \mathbf{i} \in \mathbb{Z}^2.$$

Podemos ainda descrever a dependência entre os eventos $\{X_{\mathbf{A}} \leq x_{\mathbf{A}}\}$ e $\{X_{T_{s_j}^k(\mathbf{A})} \leq x_{T_{s_j}^k(\mathbf{A})}\}$, para cada $x \in \mathbb{R}$ e $j \in \{1, \dots, 8\}$, considerando os coeficientes $\varepsilon_1(\mathbf{A}, T_{s_j}^k(\mathbf{A}))$ ou $\varepsilon_2(\mathbf{A}, T_{s_j}^k(\mathbf{A}))$ como se segue.

Para $i \in \{1, 2\}$,

$$\mathcal{E}_i(T_s^k(\mathbf{A}), \mathbf{A}) = \begin{bmatrix} \varepsilon_i(T_{s_4}^k(\mathbf{A}), \mathbf{A}) & \varepsilon_i(T_{s_3}^k(\mathbf{A}), \mathbf{A}) & \varepsilon_i(T_{s_2}^k(\mathbf{A}), \mathbf{A}) \\ \varepsilon_i(T_{s_5}^k(\mathbf{A}), \mathbf{A}) & \varepsilon_i(T_{s_0}^k(\mathbf{A}), \mathbf{A}) & \varepsilon_i(T_{s_1}^k(\mathbf{A}), \mathbf{A}) \\ \varepsilon_i(T_{s_6}^k(\mathbf{A}), \mathbf{A}) & \varepsilon_i(T_{s_7}^k(\mathbf{A}), \mathbf{A}) & \varepsilon_i(T_{s_8}^k(\mathbf{A}), \mathbf{A}) \end{bmatrix}, \text{ com } k \geq 1. \quad (2.9)$$

Recorrendo a operações matriciais, pode generalizar-se (1.16) da seguinte forma

$$\Lambda(T_s^k(\mathbf{i}), \{\mathbf{i}\}) = \mathbf{2} - \mathcal{E}_1(T_s^k(\mathbf{i}), \{\mathbf{i}\}) = \mathbf{2} - 2\mathcal{E}_2(T_s^k(\mathbf{i}), \{\mathbf{i}\}),$$

onde $\mathbf{2}$ representa uma matriz de ordem 3 com todos os elementos iguais a 2.

A utilização das oito direções s_j , $j \in \{1, \dots, 8\}$, permite localizar de forma exata as posições dos conjuntos. Contudo, podem considerar-se quatro ou simplesmente duas direções na definição das matrizes anteriores. Por uma questão de simplicidade de escrita, será adotada esta abordagem nas próximas secções.

2.3 Coeficiente médio de dependência na cauda superior

Para quaisquer localizações \mathbf{i} e \mathbf{j} em \mathbb{Z}^2 tem-se

$$\lambda(\{\mathbf{j}\}, \{\mathbf{i}\}) = \lambda(\{s_k^m \circ s_l^n(\mathbf{i})\}, \{\mathbf{i}\})$$

para $k, l \in \{1, 3, 5, 7\}$ e $m, n \in \mathbb{N}_0$. Deste modo, os coeficientes de dependência entre \mathbf{i} e os seus vizinhos podem ser obtidos a partir de matrizes de dependência de ordem superior, como mostraremos no que se segue.

Para além dos elementos da matriz $\Lambda(T_s^k(\mathbf{i}), \mathbf{i})$, $k \geq 1$, iremos considerar coeficientes de dependência resultantes de mais composições $s_k^m \circ s_l^n$, com $k \neq l$, obtendo-se assim uma nova matriz de dependência,

$$\Lambda^k(\mathbf{i}) = [\lambda_{p,q}^k(\mathbf{i})]_{\substack{p \in \{1, \dots, 2k+1\} \\ q \in \{1, \dots, 2k+1\}}} \quad (2.10)$$

cujos elementos $\lambda_{p,q}^k(\mathbf{i})$ são tais que

$$\lambda_{k+1, k+1}^k(\mathbf{i}) = 1$$

2.3. COEFICIENTE MÉDIO DE DEPENDÊNCIA NA CAUDA SUPERIOR

e, para $l \in \{0, \dots, k\}$ e $m \in \{0, \dots, k\}$,

$$\begin{aligned}\lambda_{k+1-l, k+1+m}^k(\mathbf{i}) &= \lambda(\{s_3^l \circ s_1^m(\mathbf{i})\}, \{\mathbf{i}\}), \\ \lambda_{k+1-l, k+1-m}^k(\mathbf{i}) &= \lambda(\{s_3^l \circ s_5^m(\mathbf{i})\}, \{\mathbf{i}\}), \\ \lambda_{k+1+l, k+1-m}^k(\mathbf{i}) &= \lambda(\{s_5^m \circ s_7^l(\mathbf{i})\}, \{\mathbf{i}\}), \\ \lambda_{k+1+l, k+1+m}^k(\mathbf{i}) &= \lambda(\{s_7^l \circ s_1^m(\mathbf{i})\}, \{\mathbf{i}\}),\end{aligned}$$

ou equivalentemente

$$\lambda_{p,q}^k(\mathbf{i}) = \begin{cases} \lambda(\{s_3^{k+1-p} \circ s_5^{k+1-q}(\mathbf{i})\}, \{\mathbf{i}\}) & \text{se } 1 \leq p \leq k+1, \quad 1 \leq q \leq k+1 \\ \lambda(\{s_5^{k+1-q} \circ s_7^{p-k-1}(\mathbf{i})\}, \{\mathbf{i}\}) & \text{se } k+1 < p \leq 2k+1, \quad 1 \leq q \leq k+1 \\ \lambda(\{s_3^{k+1-p} \circ s_1^{q-k-1}(\mathbf{i})\}, \{\mathbf{i}\}) & \text{se } 1 \leq p \leq k+1, \quad k+1 < q \leq 2k+1 \\ \lambda(\{s_7^{p-k-1} \circ s_1^{q-k-1}(\mathbf{i})\}, \{\mathbf{i}\}) & \text{se } k+1 < p \leq 2k+1, \quad k+1 < q \leq 2k+1 \end{cases}.$$

Tem-se

$$\Lambda^1(\mathbf{i}) = \Lambda(T_s^1(\mathbf{i}), \mathbf{i}),$$

e pode-se extrair $\Lambda(T_s^n(\mathbf{i}), \mathbf{i})$, para $n \leq k$, a partir de $\Lambda^k(\mathbf{i})$, considerando as suas diagonais, a linha $\lambda_{k+1,q}^k(\mathbf{i})$, $q \in \{1, \dots, 2k+1\}$ e a coluna $\lambda_{p,k+1}^k(\mathbf{i})$, $p \in \{1, \dots, 2k+1\}$.

Como poderemos ver na próxima secção, mesmo em exemplos simples, determinar $\lambda(\mathbf{A}, \mathbf{B})$ pode revelar-se uma tarefa árdua quando os conjuntos \mathbf{A} e \mathbf{B} têm mais do que uma localização. Tal motivou a introdução de uma nova medida de dependência, a média de coeficientes de dependência na cauda obtidos a partir das matrizes $\Lambda^k(\mathbf{i})$, $k \geq 1$.

Definição 2.1 Para quaisquer conjuntos de localizações \mathbf{A} e \mathbf{B} em \mathbb{Z}^2 , o coeficiente médio de dependência na cauda, $\bar{\lambda}(\mathbf{B}, \mathbf{A})$, define-se por

$$\bar{\lambda}(\mathbf{B}, \mathbf{A}) = \frac{1}{|\mathbf{B}|} \sum_{\mathbf{j} \in \mathbf{B}} \lambda(\{\mathbf{j}\}, \mathbf{A}). \quad (2.11)$$

No que se segue iremos considerar o caso particular em que \mathbf{B} é um conjunto de vizinhos de $\mathbf{A} = \{\mathbf{i}\}$.

Para cada $n \geq 1$, o conjunto de vizinhos de uma localização \mathbf{i} a passo n é definido por

$$\mathcal{V}^n(\mathbf{i}) = \{\mathbf{j} = (j_1, j_2) \in \mathbb{Z}^2 : |j_1 - i_1| \vee |j_2 - i_2| = n\},$$

ou de forma equivalente,

$$\mathcal{V}^n(\mathbf{i}) = \{((-1)^u s_1^n \circ (-1)^v s_3^n)(\mathbf{i}), ((-1)^u s_3^n \circ (-1)^v s_1^n)(\mathbf{i}) : u, v \in \{1, 2\}, r, m \in \{0, \dots, n\}\},$$

onde $-s_1^n = s_5^n$, $-s_3^n = s_7^n$.

Atendendo a que a matriz $\Lambda^k(\mathbf{i})$ pode ser obtida a partir de $\Lambda^{k-1}(\mathbf{i})$, acrescentando as linhas e as colunas correspondentes a $\lambda(\{\mathbf{j}\}, \{\mathbf{i}\})$ com $\mathbf{j} \in \mathcal{V}^k(\mathbf{i})$, então podemos utilizá-la na determinação do coeficiente médio de dependência na cauda entre \mathbf{i} e os seus vizinhos a passo k .

Proposição 2.2 *Para qualquer localização $\mathbf{i} \in \mathbb{Z}^2$ e $k \geq 1$, tem-se*

1.

$$\bar{\lambda}(\mathcal{V}^k(\mathbf{i}), \{\mathbf{i}\}) = (8k)^{-1} \left(\sum_{q=1}^{2k+1} (\lambda_{1,q}^k(\mathbf{i}) + \lambda_{2k+1,q}^k(\mathbf{i})) + \sum_{p=2}^{2k} (\lambda_{p,1}^k(\mathbf{i}) + \lambda_{p,2k+1}^k(\mathbf{i})) \right); \quad (2.12)$$

2.

$$\bar{\lambda} \left(\bigcup_{n=1}^k \mathcal{V}^n(\mathbf{i}), \{\mathbf{i}\} \right) = (4k(k+1))^{-1} \left(\sum_{p,q} \lambda_{p,q}^k(\mathbf{i}) - 1 \right).$$

Demonstração: Estes resultados são consequência imediata das definições de $\Lambda^k(\mathbf{i})$ e de $\bar{\lambda}(\mathbf{B}, \mathbf{A})$. \square

Em geral, para obter $\bar{\lambda}(\mathcal{V}^t(\mathbf{i}), \{\mathbf{m}\})$ com $\mathbf{m} \in \mathcal{V}^k(\mathbf{i})$ e $t > k \geq 1$, temos de conhecer $\Lambda^s(\mathbf{m})$, onde

$$s = \min\{n : \mathcal{V}^t(\mathbf{i}) \subseteq \mathcal{V}^n(\mathbf{m})\}.$$

Contudo, se $\{X_{\mathcal{V}^t(\mathbf{i})} > x_{\mathcal{V}^t(\mathbf{i})}\}$ e $\{X_{\mathbf{i}} > x\}$ são acontecimentos condicionalmente independentes relativamente a $F \in \sigma(X_{\mathbf{m}})$, onde $\sigma(X_{\mathbf{m}})$ denota a σ -álgebra gerada pela v.a. $X_{\mathbf{m}}$, ainda podemos utilizar a matriz $\Lambda^t(\mathbf{i})$, como se mostra na proposição seguinte.

Proposição 2.3 *Seja $\mathbf{i} \in \mathbb{Z}^2$ e $t > k \geq 1$. Se, para alguma escolha de $E \in \sigma(X_{\mathbf{j}}, \mathbf{j} \in \mathcal{V}^t(\mathbf{i}))$, $F \in \sigma(X_{\mathbf{j}}, \mathbf{j} \in \mathcal{V}^k(\mathbf{i}))$ e $G \in \sigma(X_{\mathbf{i}})$, E e G são condicionalmente independentes relativamente a F , então*

1. para cada $\mathbf{B} \subseteq \mathcal{V}^t(\mathbf{i})$ e $\mathbf{A} \subseteq \mathcal{V}^k(\mathbf{i})$

$$\lambda(\mathbf{B}, \{\mathbf{i}\}) = \lambda(\mathbf{B}, \mathbf{A})\lambda(\mathbf{A}, \{\mathbf{i}\});$$

2. para cada $\mathbf{m} \in \mathcal{V}^k(\mathbf{i})$,

$$\begin{aligned} \bar{\lambda}(\mathcal{V}^t(\mathbf{i}), \{\mathbf{m}\}) &= \frac{\bar{\lambda}(\mathcal{V}^t(\mathbf{i}), \{\mathbf{i}\})}{\lambda(\{\mathbf{m}\}, \{\mathbf{i}\})} \\ &= \frac{(8t)^{-1} \left(\sum_{q=1}^{2t+1} (\lambda_{1,q}^t(\mathbf{i}) + \lambda_{2t+1,q}^t(\mathbf{i})) + \sum_{p=2}^{2t} (\lambda_{p,1}^t(\mathbf{i}) + \lambda_{p,2t+1}^t(\mathbf{i})) \right)}{\lambda(\{\mathbf{m}\}, \{\mathbf{i}\})}, \end{aligned}$$

2.3. COEFICIENTE MÉDIO DE DEPENDÊNCIA NA CAUDA SUPERIOR

desde que $\lambda(\{\mathbf{m}\}, \{\mathbf{i}\}) \neq 0$.

Demonstração:

1. Para $\mathbf{B} \subseteq \mathcal{V}^t(\mathbf{i})$ e $\mathbf{A} \subseteq \mathcal{V}^k(\mathbf{i})$ tem-se

$$\begin{aligned}
\lambda(\mathbf{B}, \{\mathbf{i}\}) &= \lim_{x \rightarrow x^F} P \left(\bigcap_{j \in \mathbf{B}} \{X_j > x\} \mid \{X_i > x\} \right) \\
&= \lim_{x \rightarrow x^F} P \left(\bigcap_{j \in \mathbf{B}} \{X_j > x\}, \bigcap_{j \in \mathbf{A}} \{X_j > x\} \mid \{X_i > x\} \right) + \\
&+ \lim_{x \rightarrow x^F} P \left(\bigcap_{j \in \mathbf{B}} \{X_j > x\}, \bigcup_{j \in \mathbf{A}} \{X_j \leq x\} \mid \{X_i > x\} \right) \\
&= \lim_{x \rightarrow x^F} P \left(\bigcap_{j \in \mathbf{B}} \{X_j > x\} \mid \bigcap_{j \in \mathbf{A}} \{X_j > x\}, \{X_i > x\} \right) \times \\
&\times P \left(\bigcap_{j \in \mathbf{A}} \{X_j > x\} \mid \{X_i > x\} \right) + \\
&+ \lim_{x \rightarrow x^F} P \left(\bigcap_{j \in \mathbf{B}} \{X_j > x\} \mid \bigcup_{j \in \mathbf{A}} \{X_j \leq x\}, \{X_i > x\} \right) \times \\
&\times P \left(\bigcup_{j \in \mathbf{A}} \{X_j \leq x\} \mid \{X_i > x\} \right) \\
&= \lim_{x \rightarrow x^F} P \left(\bigcap_{j \in \mathbf{B}} \{X_j > x\} \mid \bigcap_{j \in \mathbf{A}} \{X_j > x\} \right) \times P \left(\bigcap_{j \in \mathbf{A}} \{X_j > x\} \mid \{X_i > x\} \right) + \\
&+ \lim_{x \rightarrow x^F} P \left(\bigcap_{j \in \mathbf{B}} \{X_j > x\} \mid \bigcup_{j \in \mathbf{A}} \{X_j \leq x\} \right) \times P \left(\bigcup_{j \in \mathbf{A}} \{X_j \leq x\} \mid \{X_i > x\} \right) \\
&= \lambda(\mathbf{B}, \mathbf{A}) \lambda(\mathbf{A}, \{\mathbf{i}\}) + \lim_{x \rightarrow x^F} P \left(\bigcap_{j \in \mathbf{B}} \{X_j > x\} \mid \bigcup_{j \in \mathbf{A}} \{X_j \leq x\} \right) \times \\
&\times \left(1 - P \left(\bigcap_{j \in \mathbf{A}} \{X_j > x\} \mid \{X_i > x\} \right) \right) \\
&= \lambda(\mathbf{B}, \mathbf{A}) \lambda(\mathbf{A}, \{\mathbf{i}\}) + \\
&+ \lim_{x \rightarrow x^F} P \left(\bigcap_{j \in \mathbf{B}} \{X_j > x\} \mid \bigcup_{j \in \mathbf{A}} \{X_j \leq x\} \right) (1 - \lambda(\mathbf{A}, \{\mathbf{i}\})) \\
&= \lambda(\mathbf{B}, \mathbf{A}) \lambda(\mathbf{A}, \{\mathbf{i}\}).
\end{aligned}$$

2. O resultado obtém-se de (2.12), atendendo a que para $\mathbf{j} \in \mathcal{V}^t(\mathbf{i})$ e $\mathbf{m} \in \mathcal{V}^k(\mathbf{i})$ podemos escrever

$$\lambda(\{\mathbf{j}\}, \{\mathbf{i}\}) = \lambda(\{\mathbf{j}\}, \{\mathbf{m}\}) \lambda(\{\mathbf{m}\}, \{\mathbf{i}\})$$

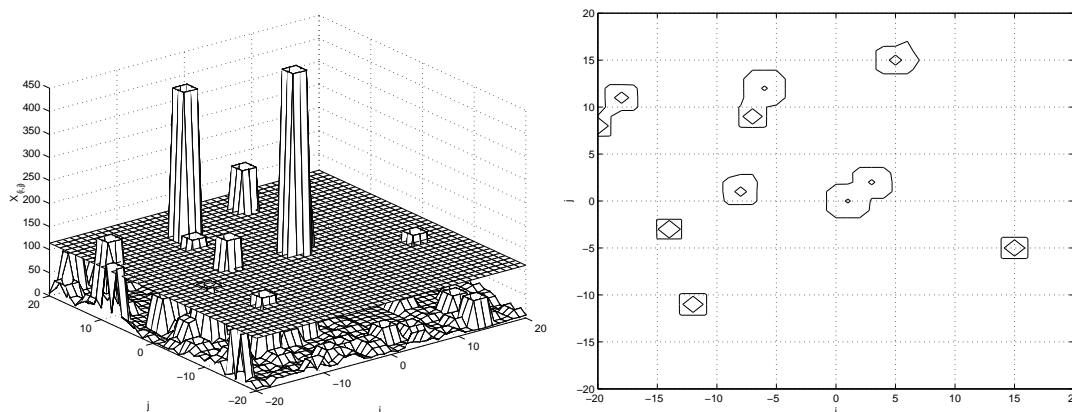


Figura 2.1: Simulação do campo aleatório \mathbf{X} definido em (2.13) (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 114.3488$ (direita).

e

$$\bar{\lambda}(\mathcal{V}^t(\mathbf{i}), \{\mathbf{i}\}) = \bar{\lambda}(\mathcal{V}^t(\mathbf{i}), \{\mathbf{m}\})\lambda(\{\mathbf{m}\}, \{\mathbf{i}\}).$$

□

Os elementos da matriz $\Lambda^k(\mathbf{i})$ relacionam-se com os elementos da matriz $\Lambda^t(\mathbf{i})$, $t > k \geq 1$, do seguinte modo

$$\begin{aligned} & \sum_{q=1}^{2k+1} (\lambda_{1,q}^k(\mathbf{i}) + \lambda_{2k+1,q}^k(\mathbf{i})) + \sum_{p=2}^{2k} (\lambda_{p,1}^k(\mathbf{i}) + \lambda_{p,2k+1}^k(\mathbf{i})) \\ &= \sum_{q=1+t-k}^{t+k+1} (\lambda_{1+t-k,q}^t(\mathbf{i}) + \lambda_{t+k+1,q}^t(\mathbf{i})) + \sum_{p=2+t-k}^{t+k} (\lambda_{p,1+t-k}^t(\mathbf{i}) + \lambda_{p,t+k+1}^t(\mathbf{i})). \end{aligned}$$

2.4 Exemplos

Nesta secção aplicamos as matrizes de dependência, introduzidas anteriormente, a dois campos aleatórios. No primeiro exemplo considera-se um campo aleatório 1-dependente max-estável e anisotrópico. O segundo exemplo é o processo Gaussiano extremal definido por Schlather (2002).

Exemplo 2.1 Seja $\mathbf{Y} = \{Y_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{Z}^2}$ um campo aleatório i.i.d. com f.d. marginal univariada de Fréchet unitária. Considere-se o campo aleatório 1-dependente, max-estável e anisotrópico $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{Z}^2}$ (Figura 2.1) definido do seguinte modo

$$X_{\mathbf{i}} = \max\{Y_{s_j(\mathbf{i})}, j \in \{1, \dots, 8\}\}, \mathbf{i} \in \mathbb{Z}^2. \quad (2.13)$$

A propensão para a ocorrência simultânea de valores elevados de \mathbf{X} localizados em $\mathbf{A} = \{\mathbf{i}, \mathbf{i} + \mathbf{1}\}$ e na região formada pelas translações de \mathbf{A} a passo 1, segundo

cada uma das direções, é dada por

$$\Lambda(T_s^1(\mathbf{A}), \mathbf{A}) = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/2 \end{bmatrix},$$

obtendo-se a independência em todas as direções no passo 2,

$$\Lambda(T_s^2(\mathbf{A}), \mathbf{A}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

A partir das matrizes dos coeficientes de dependência na cauda superior entre $\mathbf{A} = \{\mathbf{i}\}$ e as suas translações, a passos 1, 2 e 3, segundo todas as direções,

$$\begin{aligned} \Lambda(T_s^1(\{\mathbf{i}\}), \{\mathbf{i}\}) &= \begin{bmatrix} 2/8 & 4/8 & 2/8 \\ 4/8 & 1 & 4/8 \\ 2/8 & 4/8 & 2/8 \end{bmatrix} & \Lambda(T_s^2(\{\mathbf{i}\}), \{\mathbf{i}\}) &= \begin{bmatrix} 1/8 & 3/8 & 1/8 \\ 3/8 & 1 & 3/8 \\ 1/8 & 3/8 & 1/8 \end{bmatrix} \\ \Lambda(T_s^3(\{\mathbf{i}\}), \{\mathbf{i}\}) &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, & & (2.14) \end{aligned}$$

constata-se que a independência é obtida em todas as direções, no passo 3.

Combinando a informação contida nas matrizes (2.14) com os coeficientes de dependência na cauda superior em outras direções, obtém-se a seguinte matriz estendida

$$\Lambda^3(\mathbf{i}) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/8 & 2/8 & 3/8 & 2/8 & 1/8 & 0 \\ 0 & 2/8 & 2/8 & 4/8 & 2/8 & 2/8 & 0 \\ 0 & 3/8 & 4/8 & 1 & 4/8 & 3/8 & 0 \\ 0 & 2/8 & 2/8 & 4/8 & 2/8 & 2/8 & 0 \\ 0 & 1/8 & 2/8 & 3/8 & 2/8 & 1/8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (2.15)$$

Na Figura 2.2 ilustramos, através de uma escala de cinzas, a dependência na cauda bivariada de \mathbf{X} , de acordo com a matriz (2.15). Os tons mais claros correspondem a menor dependência e os mais escuros a maior dependência, sendo a dependência total representada a preto e a independência a branco.

Aplicando a Proposição 2.2 obtemos $\bar{\lambda}(\mathcal{V}^k(\mathbf{i}), \{\mathbf{i}\})$, $k \leq 3$, a partir dos elementos

3	3	3	3	3	3	3
3	2	2	2	2	2	3
3	2	1	1	1	2	3
3	2	1	1	1	2	3
3	2	1	1	1	2	3
3	2	2	2	2	2	3
3	3	3	3	3	3	3

Figura 2.2: Representação, numa escala de cinzas, dos valores de dependência da matriz $\Lambda^3(\mathbf{i})$.

de $\Lambda^3(\mathbf{i})$,

$$\bar{\lambda}(\mathcal{V}^1(\mathbf{i}), \{\mathbf{i}\}) = \frac{3}{8}, \quad \bar{\lambda}(\mathcal{V}^2(\mathbf{i}), \{\mathbf{i}\}) = \frac{1}{4}, \quad \bar{\lambda}(\mathcal{V}^k(\mathbf{i}), \{\mathbf{i}\}) = 0, \quad \text{para } k \geq 3.$$

Saliente-se que $\bar{\lambda}(\mathcal{V}^t(\mathbf{i}), \{\mathbf{m}\})$ com $\mathbf{m} \in \mathcal{V}^k(\mathbf{i})$, $t > k$, não pode ser obtido da Proposição 2.3 para $k \geq 3$ dado que $\lambda(\{\mathbf{m}\}, \{\mathbf{i}\}) = 0$. Deste modo, para calcular $\bar{\lambda}(\mathcal{V}^3(\mathbf{i}), \{s_3^2 \circ s_1(\mathbf{i})\})$ podemos utilizar diretamente os valores em $\Lambda^5(\mathbf{i})$ ou simplesmente os valores em $\Lambda^3(\mathbf{i})$ uma vez que $\lambda(\mathcal{V}^k(\mathbf{i}), \{\mathbf{i}\}) = 0$, $k \geq 3$, obtendo-se

$$\begin{aligned} \bar{\lambda}(\mathcal{V}^3(\mathbf{i}), \{s_3^2 \circ s_1(\mathbf{i})\}) &= \frac{1}{|\mathcal{V}^3(\mathbf{i})|} \sum_{\mathbf{j} \in \mathcal{V}^3(\mathbf{i})} \lambda(\{\mathbf{j}\}, \{s_3^2 \circ s_1(\mathbf{i})\}) \\ &= \frac{1}{24} \left(5 \times \frac{2}{8} + \frac{4}{8} + \frac{3}{8} + \frac{1}{8} \right) = \frac{3}{32}. \end{aligned}$$

Neste caso, as matrizes dos coeficientes extremais bivariados, definidas em (2.9), são dadas por

$$\mathcal{E}_2(T_s^1(\{\mathbf{i}\}), \{\mathbf{i}\}) = \begin{bmatrix} 14/16 & 12/16 & 14/16 \\ 12/16 & 1/2 & 12/16 \\ 14/16 & 12/16 & 14/16 \end{bmatrix}, \quad \mathcal{E}_2(T_s^2(\{\mathbf{i}\}), \{\mathbf{i}\}) = \begin{bmatrix} 15/16 & 13/16 & 15/16 \\ 13/16 & 1/2 & 13/16 \\ 15/16 & 13/16 & 15/16 \end{bmatrix}$$

$$\mathcal{E}_2(T_s^3(\{\mathbf{i}\}), \{\mathbf{i}\}) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1/2 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

e permitem quantificar o grau de dependência entre os eventos $\{X_{\mathbf{i}} \leq x_{\mathbf{i}}\}$ e $\{X_{T_{s_j}^k(\mathbf{i})} \leq x_{T_{s_j}^k(\mathbf{i})}\}$, $j \in \{1, \dots, 8\}$ para cada $x \in \mathbb{R}$ e $k \in \{1, 2, 3\}$. Refira-se que a total dependência corresponde o valor 1/2 e o valor 1 à independência. À medida que k aumenta o valor dos oito elementos da matriz, correspondentes às oito direções, aumenta, o que corresponde a uma diminuição da dependência.

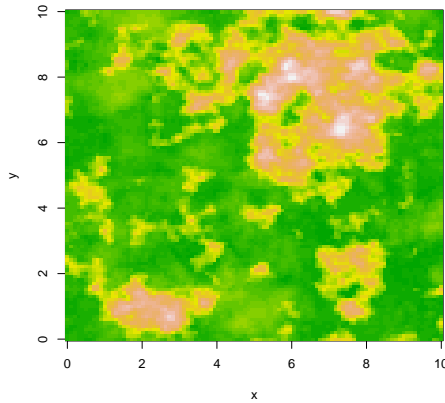


Figura 2.3: Modelo de Schlather com $\rho(\|\mathbf{i} - \mathbf{j}\|) = \exp(-\sqrt{(i_1 - j_1)^2 + (i_2 - j_2)^2})$.

Exemplo 2.2 Consideremos a classe de distribuições bivariadas, associada ao processo Gaussiano extremal, \mathbf{X} , de Schlather (2002), definida por

$$P(X_{\mathbf{i}} \leq x, X_{\mathbf{j}} \leq y) = \exp \left[-\frac{1}{2} \left(\frac{1}{x} + \frac{1}{y} \right) \left(1 + \sqrt{1 - \frac{2(\rho(\|\mathbf{i} - \mathbf{j}\|) + 1)xy}{(x+y)^2}} \right) \right],$$

onde $x, y \in \mathbb{R}$ e $\rho(\cdot)$ representa a função de covariância do processo Gaussiano subjacente.

No que se segue utilizamos a função de covariância definida por

$$\rho(\|\mathbf{i} - \mathbf{j}\|) = \exp(-\|\mathbf{i} - \mathbf{j}\|),$$

com $\mathbf{i} = (i_1, i_2)$ e $\mathbf{j} = (j_1, j_2)$, onde $\|\mathbf{i} - \mathbf{j}\| = \sqrt{(i_1 - j_1)^2 + (i_2 - j_2)^2}$ (Figura 2.3).

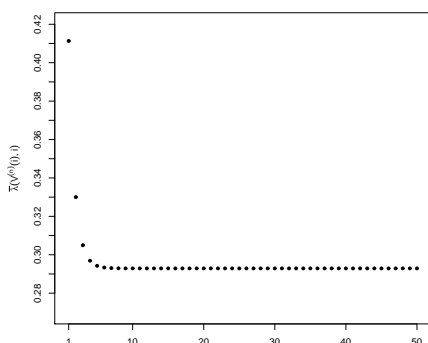
O coeficiente extremal bivariado é dado por

$$\varepsilon(\{\mathbf{i}, \mathbf{j}\}) = 1 + \sqrt{\frac{1 - \rho(\|\mathbf{i} - \mathbf{j}\|)}{2}}$$

e atendendo a que o campo \mathbf{X} é isotrópico, para a determinação das matrizes $\Lambda(T_s^k(\{\mathbf{i}\}), \{\mathbf{i}\})$ apenas precisamos de determinar os seguintes coeficientes

$$\lambda(T_{s_j}^k(\{\mathbf{i}\}), \{\mathbf{i}\}) = 1 - \sqrt{\frac{1 - \rho(\|T_{s_j}^k(\{\mathbf{i}\}) - \mathbf{i}\|)}{2}}, \quad j \in \{1, 2\}. \quad (2.16)$$

Considerando $j \in \{1, \dots, 8\}$ em (2.16) determinaram-se os valores necessários para definir a matriz (2.17) de coeficientes de dependência na cauda superior, descrita na Secção 2.3. Os elementos desta matriz quantificam a dependência na cauda entre $X_{\mathbf{i}}$ e os seus vizinhos até ao passo 3 e mostram, tal como se esperava, que à


 Figura 2.4: Dependência entre \mathbf{i} e $\mathcal{V}^n(\mathbf{i})$, $n \geq 1$, no modelo de Schlather.

medida que k aumenta a dependência diminui.

$$\Lambda^3(\mathbf{i}) = \begin{bmatrix} 0.298 & 0.303 & 0.308 & 0.312 & 0.308 & 0.303 & 0.298 \\ 0.303 & 0.314 & 0.332 & 0.343 & 0.332 & 0.314 & 0.303 \\ 0.308 & 0.332 & 0.385 & 0.438 & 0.385 & 0.332 & 0.308 \\ 0.312 & 0.343 & 0.438 & 1 & 0.438 & 0.343 & 0.312 \\ 0.308 & 0.332 & 0.385 & 0.438 & 0.385 & 0.332 & 0.308 \\ 0.303 & 0.314 & 0.332 & 0.343 & 0.332 & 0.314 & 0.303 \\ 0.298 & 0.303 & 0.308 & 0.312 & 0.308 & 0.303 & 0.298 \end{bmatrix}. \quad (2.17)$$

O coeficiente médio de dependência na cauda é dado, para $n \geq 1$, por

$$\begin{aligned} \bar{\lambda}(\mathcal{V}^n(\mathbf{i}), \{\mathbf{i}\}) &= \frac{1}{|\mathcal{V}^n(\mathbf{i})|} \sum_{\mathbf{j} \in \mathcal{V}^n(\mathbf{i})} \lambda(\{\mathbf{i}\}, \{\mathbf{j}\}) \\ &= \frac{1}{8n} \sum_{\mathbf{j} \in \mathcal{V}^n(\mathbf{i})} \left(1 - \sqrt{\frac{1 - \rho(\|\mathbf{i} - \mathbf{j}\|)}{2}} \right) \\ &= 1 - \frac{1}{8n} \left(4\sqrt{\frac{1 - \exp(-n)}{2}} + 4\sqrt{\frac{1 - \exp(-n\sqrt{2})}{2}} + \sum_{k=1}^{n-1} 8\sqrt{\frac{1 - \exp(-\sqrt{n^2 + k^2})}{2}} \right). \end{aligned}$$

A partir da sua representação gráfica (Figura 2.4), para alguns valores de $n \geq 1$, constatamos, tal como anteriormente, que quando n aumenta $\bar{\lambda}(\mathcal{V}^n(\mathbf{i}), \mathbf{i})$ diminui.

Notemos que $\rho(\|\mathbf{i} - \mathbf{j}\|) \rightarrow 0$ quando $\|\mathbf{i} - \mathbf{j}\| \rightarrow +\infty$, e por conseguinte $\varepsilon(\{\mathbf{i}, \mathbf{j}\}) \leq 1 + \sqrt{1/2}$, logo a independência nunca surgirá.

- Calcular $S(\mathbf{J})$:
 1. Determinar $\varepsilon = \frac{\log D_G(\mathbf{u}^{(|\mathbf{J}|)})}{\log u}$.
 2. Determinar $S(\mathbf{J}) = |\mathbf{J}| + (-1)^{|\mathbf{J}|+1} \times \varepsilon$.
 - (a) Para $i = 2$ até $|\mathbf{J}| - 1$
 - i. Determinar $Nset = \binom{|\mathbf{J}|}{i}$
 - ii. Para $j = 1$ até $Nset$
 - A. Determinar $\varepsilon = \frac{\log D_G(\mathbf{u}^{(i)})}{\log u}$.
 - B. Atualizar $S(\mathbf{J}) : S(\mathbf{J}) + (-1)^{i+1} \times \varepsilon$.
- Determinar $\lambda(\mathbf{B}, \mathbf{A}) = \frac{S(\mathbf{A} \cup \mathbf{B})}{S(\mathbf{A})}$.

Figura 2.5: Algoritmo que devolve o coeficiente de dependência na cauda superior $\lambda(\mathbf{B}, \mathbf{A})$ dados dois conjuntos de localizações \mathbf{A} e \mathbf{B} em \mathbb{Z}^2 e uma função de dependência de valores extremos $|\mathbf{A} \cup \mathbf{B}|$ -dimensional, $D_G(\mathbf{u}^{(|\mathbf{A} \cup \mathbf{B}|)})$.

2.5 Algoritmo para determinar os coeficientes de dependência

Os exemplos apresentados evidenciam que o cálculo dos coeficientes de dependência na cauda superior $\lambda(\mathbf{B}, \mathbf{A})$ é um processo trabalhoso quando os conjuntos \mathbf{A} e \mathbf{B} têm mais do que uma localização. Considerando a relação (2.1) com

$$\varepsilon(\mathbf{J}) = \frac{\log D_G(\mathbf{u}^{(|\mathbf{J}|)})}{\log u}, \quad \mathbf{J} \subseteq \mathbb{Z}^2,$$

onde D_G denota a função de dependência associada à f.d. G do vetor aleatório formado pelas variáveis aleatórias X_i , $i \in \mathbf{J}$, e $\mathbf{u}^{(|\mathbf{J}|)}$ um vetor de dimensão $|\mathbf{J}|$ com todas as componentes iguais a $u \in [0, 1]$, propõe-se um algoritmo simples (Figura 2.5) para determinar o valor de $\lambda(\mathbf{A}, \mathbf{B})$, dados os conjuntos \mathbf{A} e \mathbf{B} de \mathbb{Z}^2 .

O algoritmo foi implementado no *Mathematica 7*, por permitir cálculo simbólico e, como tal, permitir obter a expressão algébrica dos coeficientes. Também foi neste programa que se introduziram as instruções necessárias para determinar os coeficientes $\varepsilon_1(\mathbf{A}, \mathbf{B})$, $\varepsilon_2(\mathbf{A}, \mathbf{B})$ e $\bar{\lambda}(\mathbf{A}, \mathbf{B})$, apresentados anteriormente. Refira-se que para conjuntos de localizações de grande dimensão a aplicação do algoritmo exigirá maior tempo de computação, dado o aumento da complexidade.

Aplicou-se o algoritmo ao caso particular de dois conjuntos de localizações, \mathbf{A} e \mathbf{B} , e considerando as funções de dependência multivariadas de valores extremos de

Gumbel e da família MM3 de Joe (1997, p. 164).

Exemplo 2.3 Considere-se o vetor aleatório formado pelas variáveis aleatórias X_i indexadas nos seguintes conjuntos \mathbf{A} e \mathbf{B} de \mathbb{Z}^2

$$\mathbf{A} = \{(1, 1), (1, 2), (1, 3), (2, 2), (2, 1), (2, 3)\},$$

$$\mathbf{B} = \{(3, 1), (3, 2), (3, 3), (3, 4), (4, 1), (4, 2), (4, 3), (4, 4)\},$$

com função de dependência de Gumbel (Figura 2.6)

$$D_G(u_1, \dots, u_{|\mathbf{A} \cup \mathbf{B}|}) = \exp \left(- \left(\sum_{j=1}^{|\mathbf{A} \cup \mathbf{B}|} (-\log u_j)^\delta \right)^{\frac{1}{\delta}} \right), \quad u_j \in [0, 1], \quad \delta \geq 1, \quad (2.18)$$

onde $\delta = 1$ corresponde à situação de independência e $\delta \rightarrow +\infty$ à situação de total dependência.

Para $\delta = 2$ obtém-se

$$\begin{aligned} \varepsilon(\mathbf{A}) &= 2.4495, & \varepsilon(\mathbf{B}) &= 2.8284, & \lambda(\mathbf{A}, \mathbf{B}) &= 0.8345, \\ \varepsilon_1(\mathbf{A}, \mathbf{B}) &= 1.5275, & \varepsilon_2(\mathbf{A}, \mathbf{B}) &= 0.7089, & \bar{\lambda}(\mathbf{A}, \mathbf{B}) &= 0.9628. \end{aligned}$$

Exemplo 2.4 Neste exemplo consideramos que a função de dependência do vetor aleatório formado pelas variáveis X_i com $\mathbf{i} \in \mathbf{A} \cup \mathbf{B}$,

$$\mathbf{A} = \{(1, 1), (1, 2)\} \text{ e } \mathbf{B} = \{(2, 1), (2, 2), (2, 3)\},$$

pertence à família MM3 das funções de dependência de valores extremos, definida em Joe (1997) por

$$D_G(u_1, \dots, u_n) = \exp \left(- \left(\sum_{j=1}^n \left(z_j^\theta - \sum_{i < j} (p_i^{-\delta_{ij}} z_i^{-\theta \delta_{ij}} + p_j^{-\delta_{ij}} z_j^{-\theta \delta_{ij}})^{-\frac{1}{\delta_{ij}}} \right) \right)^{\frac{1}{\theta}} \right), \quad (2.19)$$

onde $z_j = -\log u_j$, $p_j = (\nu_j + n - 1)^{-1}$, $\nu_j \geq 0$, $j = 1, \dots, n$, $\delta_{ij} > 0$, $i < j$ e $\theta > 0$.

Na Figura 2.6 apresentamos uma simulação desta função para $n = 2$, $p = (1/92, 1/95)$, $\delta_{12} = 1$ e $\theta = 2$.

Tomando $\theta = 2$, $p = (1/92, 1/95, 1/51, 1/13, 1/78)$ e

$$\delta = \begin{bmatrix} 0 & 1 & 94 & 16 & 44 \\ 0 & 0 & 60 & 32 & 40 \\ 0 & 0 & 0 & 72 & 43 \\ 0 & 0 & 0 & 0 & 29 \end{bmatrix}$$

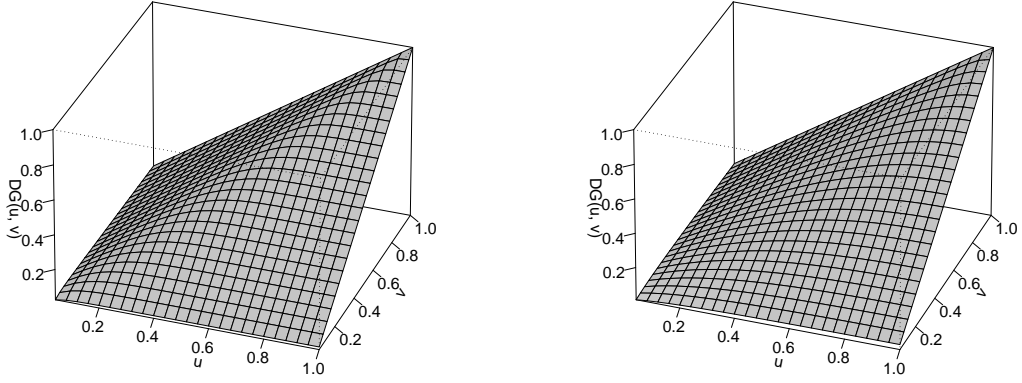


Figura 2.6: Simulação da função (2.18) com $\delta = 2$ (esquerda) e da função (2.19) com $p = (1/92, 1/95)$, $\delta_{12} = 1$ e $\theta = 2$ (direita).

obtém-se

$$\varepsilon(\mathbf{A}) = \left(2 - (p_1^{-\delta_{12}} + p_2^{-\delta_{12}})^{\frac{-1}{\delta_{12}}}\right)^{\frac{1}{\theta}} = 1.4123,$$

$$\begin{aligned} \varepsilon(\mathbf{B}) &= \left(3 - (p_1^{-\delta_{12}} + p_2^{-\delta_{12}})^{\frac{-1}{\delta_{12}}} - (p_1^{-\delta_{13}} + p_3^{-\delta_{13}})^{\frac{-1}{\delta_{13}}} - (p_2^{-\delta_{23}} + p_3^{-\delta_{23}})^{\frac{-1}{\delta_{23}}}\right)^{\frac{1}{\theta}} \\ &= 1.7189, \end{aligned}$$

$$\lambda(\mathbf{A}, \mathbf{B}) = 0.7079, \quad \varepsilon_1(\mathbf{A}, \mathbf{B}) = 1.5650, \quad \varepsilon_2(\mathbf{A}, \mathbf{B}) = 0.7059, \quad \bar{\lambda}(\mathbf{A}, \mathbf{B}) = 0.8357.$$

2.6 Campo aleatório M4

Os processos máximos multivariados de máximos móveis ou simplesmente processos M4, foram introduzidos por Smith e Weissman (1996) e têm sido utilizados para modelar o comportamento de valores extremos de várias séries temporais (Zhang e Smith (2004, 2010), Morales (2005), Zhang (2009) e Suveges (2009)) por se revelarem bem adaptados a esses casos.

Smith e Weissman (1996) mostraram que assumindo algumas condições gerais, as observações extremas em processos estacionários multivariados podem ser caracterizadas em termos de um processo limite max-estável e que os processos max-estáveis podem ser aproximados por um processo M4. Estes resultados são relevantes porque permitiram caracterizar o índice extremal de processos estacionários multivariados com base nos processos M4 (Teorema 2.3 em Smith e Weissman (1996)).

No que se segue vamos definir campo aleatório M4, apresentar a sua função de dependência, verificar que é max-estável e deduzir a expressão do coeficiente extre-

mal no contexto multivariado. Iremos ainda determinar as medidas de dependência apresentadas nas Secções 2.2 e 2.3, cujo cálculo será facilitado pelo algoritmo apresentado.

Seja $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{R}^2}$ um campo aleatório M4 definido por

$$X_{\mathbf{i}} = \max_{l \geq 1} \max_{-\infty < m < +\infty} a_{l\mathbf{m}\mathbf{i}} Z_{l,1-m}, \quad \mathbf{i} \in \mathbb{R}^2, \quad (2.20)$$

onde $\{Z_{l,n}\}_{l \geq 1, n \in \mathbb{Z}}$ é uma família de variáveis aleatórias independentes com f.d. F de Fréchet unitária e, para cada localização $\mathbf{i} \in \mathbb{R}^2$, $\{a_{l\mathbf{m}\mathbf{i}}\}_{l \geq 1, m \in \mathbb{Z}}$ são constantes não negativas tais que $\sum_{l=1}^{+\infty} \sum_{m=-\infty}^{+\infty} a_{l\mathbf{m}\mathbf{i}} = 1$.

As constantes $\{a_{l\mathbf{m}\mathbf{i}}\}_{l \geq 1, m \in \mathbb{Z}}$ são os parâmetros do campo, também designados padrões de assinaturas. O número de padrões é dado por l e m representa o número de assinaturas. Estas constantes são determinantes na estrutura de dependência do campo.

A f.d. do vetor $(X_{\mathbf{i}_1}, \dots, X_{\mathbf{i}_k})$, que denotaremos por $G_{\mathbf{A}}$, com $\mathbf{A} = \{\mathbf{i}_1, \dots, \mathbf{i}_k\}$ é dada por

$$\begin{aligned} G_{\mathbf{A}}(x_1, \dots, x_k) &= P\left(Z_{l,1-m} \leq \frac{x_1}{a_{l\mathbf{m}\mathbf{i}_1}}, \dots, Z_{l,1-m} \leq \frac{x_k}{a_{l\mathbf{m}\mathbf{i}_k}}, l \geq 1, -\infty < m < +\infty\right) \\ &= \prod_{l=1}^{+\infty} \prod_{m=-\infty}^{+\infty} \exp\left(-\bigvee_{j=1}^k \left(\frac{a_{l\mathbf{m}\mathbf{i}_j}}{x_j}\right)\right) \\ &= \prod_{l=1}^{+\infty} \prod_{m=-\infty}^{+\infty} \bigwedge_{j=1}^k \exp(-x_j^{-1} a_{l\mathbf{m}\mathbf{i}_j}), \quad x_j \in \mathbb{R}^+, \quad j \in \{1, \dots, k\}. \end{aligned}$$

e por conseguinte, a função de dependência associada a $G_{\mathbf{A}}$ define-se por

$$D_{G_{\mathbf{A}}}(u_{\mathbf{i}_1}, \dots, u_{\mathbf{i}_k}) = \prod_{l=1}^{+\infty} \prod_{m=-\infty}^{+\infty} \bigwedge_{j=1}^k u_{\mathbf{i}_j}^{a_{l\mathbf{m}\mathbf{i}_j}}, \quad (2.21)$$

com $u_{\mathbf{i}_j} = F(x_j) = \exp(-x_j^{-1})$, $j \in \{1, \dots, k\}$. A estrutura de dependência de $(X_{\mathbf{i}_1}, \dots, X_{\mathbf{i}_k})$, regulada pelos padrões de assinaturas $\{a_{l\mathbf{m}\mathbf{i}}\}_{l \geq 1, m \in \mathbb{Z}}$, é caracterizada por (2.21).

O campo \mathbf{X} é max-estável uma vez que para $t \in \mathbb{N}$, a função de dependência (2.21) satisfaz

$$D_{G_{\mathbf{A}}}^t(u_{\mathbf{i}_1}, \dots, u_{\mathbf{i}_k}) = D_{G_{\mathbf{A}}}(u_{\mathbf{i}_1}^t, \dots, u_{\mathbf{i}_k}^t),$$

para quaisquer localizações $\mathbf{i}_1, \dots, \mathbf{i}_k$.

Atendendo a (1.10) obtém-se a seguinte função expoente associada à MEV $G_{\mathbf{A}}$,

$$\begin{aligned} V_{\mathbf{A}}(x_1, \dots, x_k) &= -\ln D_{G_{\mathbf{A}}}(\exp(-x_1^{-1}), \dots, \exp(-x_k^{-1})) \\ &= \sum_{l=1}^{+\infty} \sum_{m=-\infty}^{+\infty} \bigvee_{j=1}^k (x_j^{-1} a_{lmi_j}). \end{aligned} \quad (2.22)$$

e por conseguinte, o coeficiente extremal do campo M4 é dado por

$$\varepsilon(\mathbf{A}) = V_{\mathbf{A}}(1, \dots, 1) = \sum_{l=1}^{+\infty} \sum_{m=-\infty}^{+\infty} \left(\bigvee_{j=1}^k a_{lmi_j} \right). \quad (2.23)$$

No que se segue vamos considerar um campo M4, $\mathbf{X} = \{X_i\}_{i \in \mathbb{Z}^2}$, cuja simulação apresentamos na Figura 2.7, com um número finito de padrões, $L = 1$, e de assinaturas, $1 \leq m \leq 2$, sendo as constantes a_{11i} e a_{12i} definidas do seguinte modo

- $a_{11i} = \frac{1}{2}$ e $a_{12i} = \frac{1}{2}$ se $i_1 > j_1 \wedge i_1$ par $\wedge j_1$ ímpar ;
- $a_{11i} = \frac{1}{4}$ e $a_{12i} = \frac{3}{4}$ se $i_1 \leq j_1 \wedge i_1$ par $\wedge j_1$ ímpar ;
- $a_{11i} = \frac{1}{3}$ e $a_{12i} = \frac{2}{3}$ se $i_1 > j_1 \wedge i_1$ ímpar $\wedge j_1$ par ;
- $a_{11i} = \frac{1}{8}$ e $a_{12i} = \frac{9}{8}$ se $i_1 \leq j_1 \wedge i_1$ ímpar $\wedge j_1$ par ;
- $a_{11i} = \frac{1}{6}$ e $a_{12i} = \frac{5}{6}$ se $i_1 > j_1 \wedge i_1$ par $\wedge j_1$ par ;
- $a_{11i} = \frac{1}{5}$ e $a_{12i} = \frac{4}{5}$ se $i_1 \leq j_1 \wedge i_1$ par $\wedge j_1$ par ;
- $a_{11i} = \frac{1}{7}$ e $a_{12i} = \frac{6}{7}$ se $i_1 > j_1 \wedge i_1$ ímpar $\wedge j_1$ ímpar ;
- $a_{11i} = \frac{4}{9}$ e $a_{12i} = \frac{5}{9}$ se $i_1 \leq j_1 \wedge i_1$ ímpar $\wedge j_1$ ímpar .

Vamos avaliar a dependência extremal entre variáveis de \mathbf{X} indexadas em $\mathbf{A} = \{(3, 2), (4, 5), (3, 5), (6, 4)\}$ e numa região vizinha de \mathbf{A} , através das matrizes de dependência na cauda superior e extremas.

A matriz $\Lambda^k(\mathbf{A})$ corresponde à generalização da matriz (2.10) a uma região de localizações, sendo os seus elementos $\lambda_{p,q}^k(\mathbf{A})$, $p, q \in \{1, \dots, 2k+1\}$ dados por

$$\lambda_{k+1, k+1}^k(\mathbf{A}) = 1$$

e para $l, m \in \{1, \dots, k\}$,

$$\lambda_{k+1-l, k+1+m}^k(\mathbf{A}) = \lambda(T_{s_3^l \circ s_1^m}^k(\mathbf{A}), \mathbf{A})$$

$$\lambda_{k+1-l, k+1-m}^k(\mathbf{A}) = \lambda(T_{s_3^l \circ s_5^m}^k(\mathbf{A}), \mathbf{A})$$

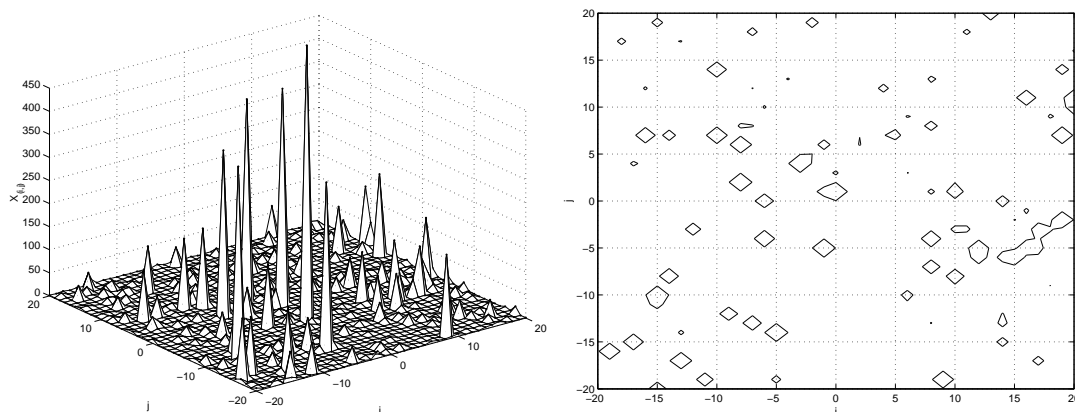


Figura 2.7: Simulação do campo aleatório M4 (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 19.3613$ (direita).

$$\lambda_{k+1+l,k+1-m}^k(\mathbf{A}) = \lambda(T_{s_5^l \circ s_7^m}^k(\mathbf{A}), \mathbf{A})$$

$$\lambda_{k+1+l,k+1+m}^k(\mathbf{A}) = \lambda(T_{s_7^l \circ s_1^m}^k(\mathbf{A}), \mathbf{A}).$$

Considerando $k = 4$ e recorrendo ao algoritmo introduzido na Secção 2.5 determinámos estes coeficientes, cujos valores apresentamos na matriz $\Lambda^4(\mathbf{A})$.

$$\Lambda^4(\mathbf{A}) = \begin{bmatrix} 49/52 & 49/52 & 49/52 & 49/52 & 49/52 & 49/52 & 49/52 & 1 & 1 \\ 49/52 & 49/52 & 49/52 & 49/52 & 49/52 & 49/52 & 45/52 & 49/52 & 49/52 \\ 49/52 & 49/52 & 49/52 & 49/52 & 49/52 & 1 & 1 & 1 & 12/13 \\ 49/52 & 49/52 & 49/52 & 49/52 & 45/52 & 45/52 & 45/52 & 81/91 & 81/91 \\ 49/52 & 49/52 & 49/52 & 1 & 1 & 1 & 12/13 & 81/91 & 81/91 \\ 49/52 & 49/52 & 45/52 & 45/52 & 45/52 & 81/91 & 45/52 & 81/91 & 81/91 \\ 49/52 & 1 & 1 & 13/18 & 12/13 & 81/91 & 81/91 & 81/91 & 81/91 \\ 45/52 & 45/52 & 45/52 & 81/91 & 81/91 & 81/91 & 81/91 & 81/91 & 81/91 \\ 1 & 13/18 & 12/13 & 81/91 & 81/91 & 81/91 & 81/91 & 81/91 & 81/91 \end{bmatrix}.$$

À semelhança da análise apresentada na Secção 2.4, na Figura 2.8 consideram-se os elementos da matriz representados numa tabela de acordo com uma escala de cinzas. Desta forma, pretendemos destacar as diferenças entre o grau de dependência na cauda das variáveis de \mathbf{X} indexadas em \mathbf{A} e das indexadas em regiões vizinhas de \mathbf{A} . Observa-se, por exemplo, que a tendência para que ocorram valores elevados de \mathbf{X} nos conjuntos vizinhos de \mathbf{A} a passo 1, sabendo que ocorreram valores elevados de \mathbf{X} em \mathbf{A} , varia com a direção que se considera. Verifica-se que existe dependência total quando se consideram os conjuntos de localizações \mathbf{A} e os seus vizinhos a passo 1 segundo as direções s_1 e s_5 . Considerando os conjuntos vizinhos de \mathbf{A} a passo k , $2 \leq k \leq 4$, também se observam diferentes graus de dependência, sendo de destacar que não existe nenhuma situação de independência.

O estudo efetuado permitiu concluir que as matrizes $\Lambda(T_s^k(\mathbf{A}), \mathbf{A})$, definidas em (2.8), quando $k > 4$ apenas apresentam duas formas distintas que dependem da paridade de k . Assim, se k é ímpar obtêm-se matrizes iguais a $\Lambda(T_s^3(\mathbf{A}), \mathbf{A})$, caso

4	4	4	4	4	4	4	4		
4	3	3	3	3	3	3	3	3	4
4	3	2	2	2					4
4	3	2	1	1	1	2	3	4	
4	3	2				2	3	4	
4	3	2	1	1	1	2	3	4	
4			2	2	2	2	3	4	
4	3	3	3	3	3	3	3	4	
	4	4	4	4	4	4	4	4	

Figura 2.8: Representação, numa escala de cinzas, dos valores de dependência da matriz $\Lambda^4(\mathbf{A})$.

contrário, obtêm-se matrizes iguais a $\Lambda(T_s^4(\mathbf{A}), \mathbf{A})$.

As matrizes de coeficientes extremais no contexto multivariado, definidas em (2.9), neste exemplo são dadas por

$$\mathcal{E}_2(T_s^1(\mathbf{A}), \mathbf{A}) = \begin{bmatrix} 95/187 & 99/191 & 99/191 \\ 115/217 & 1/2 & 1/2 \\ 99/191 & 99/191 & 171/332 \end{bmatrix};$$

$$\mathcal{E}_2(T_s^2(\mathbf{A}), \mathbf{A}) = \begin{bmatrix} 95/187 & 95/187 & 1/2 \\ 95/187 & 1/2 & 24/47 \\ 1/2 & 24/47 & 171/332 \end{bmatrix};$$

$$\mathcal{E}_2(T_s^k(\mathbf{A}), \mathbf{A}) = \begin{cases} \begin{bmatrix} 95/187 & 95/187 & 1/2 \\ 95/187 & 1/2 & 171/332 \\ 1/2 & 171/332 & 171/332 \end{bmatrix} & \text{se } k > 2 \text{ par;} \\ \begin{bmatrix} 95/187 & 95/173 & 99/191 \\ 95/187 & 1/2 & 171/332 \\ 99/191 & 171/332 & 171/332 \end{bmatrix} & \text{se } k > 2 \text{ ímpar.} \end{cases}$$

Estas matrizes de coeficientes extremais permitem quantificar a dependência entre o par de acontecimentos

$$\{X_{\mathbf{A}} \leq x_{\mathbf{A}}\} \text{ e } \{X_{T_{s_j}^k(\mathbf{A})} \leq x_{T_{s_j}^k(\mathbf{A})}\}, \quad x \in \mathbb{R}, \quad (2.24)$$

considerando $k \in \mathbb{N}$, $j \in \{1, \dots, 8\}$. Se existir independência entre um par de eventos da forma (2.24), para algum k e j , o coeficiente extremal ε_2 terá valor 1, o que nunca se verifica. Por outro lado, a dependência entre $\{X_{\mathbf{A}} \leq x_{\mathbf{A}}\}$ e cada um

CAPÍTULO 2. MATRIZES DE COEFICIENTES DE DEPENDÊNCIA

dos acontecimentos

$$\left\{ X_{T_{s_1}^1}(\mathbf{A}) \leq x_{T_{s_1}^1}(\mathbf{A}) \right\}, \left\{ X_{T_{s_2}^2}(\mathbf{A}) \leq x_{T_{s_2}^2}(\mathbf{A}) \right\}, \left\{ X_{T_{s_6}^6}(\mathbf{A}) \leq x_{T_{s_6}^6}(\mathbf{A}) \right\}$$

é significativa dado que $\varepsilon_2 = 1/2$ (limite inferior do intervalo de variação).

Capítulo 3

Madograma generalizado

3.1 Introdução

É frequente os hidrologistas agruparem dados por regiões com base nas suas características geográficas e hidrográficas, o que revela ser importante caracterizar a dependência entre extremos de duas regiões disjuntas. Um exemplo comumente referido é a dependência entre a temperatura no Golfo do México e a temperatura nas ilhas Britânicas, explicada pela chamada corrente do Golfo. Outro exemplo é o efeito da temperatura/pluviosidade em sistemas montanhosos sobre a temperatura/pluviosidade de regiões vizinhas desses sistemas.

As situações descritas anteriormente motivaram a introdução do coeficiente madograma generalizado (Fonseca *et al.* (2011)), o qual resume a dependência entre os máximos das variáveis indexadas em duas regiões disjuntas de localizações, tendo em conta a dependência inter-regiões e intra-regiões.

Após a introdução deste coeficiente, o qual generaliza o λ -madograma apresentado em Naveau *et al.* (2009) para pares de localizações, estabelecemos relações com outros coeficientes de dependência e ilustramos a sua aplicação num modelo particular, os campos M4. Com o objectivo de utilizar este coeficiente em dados reais introduzimos um estimador, para o qual provamos a normalidade assintótica e a consistência forte. Apresentamos ainda resultados de simulações que permitem avaliar o desempenho do estimador proposto e que foram obtidos através de rotinas computacionais que desenvolvemos utilizando o programa *R*.

Por fim, utilizando o trabalho desenvolvido, estudamos a dependência entre valores extremos de precipitação em regiões portuguesas com base em dados que estão disponíveis no SNIRH.

3.2 Definição e propriedades

Seja $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{R}^2}$ um campo aleatório max-estável com f.d. marginal univariada F de Fréchet unitária, isto é, $F(x) = \exp(-x^{-1})$, $x > 0$.

Para quantificar a dependência entre o máximo das variáveis de \mathbf{X} indexadas em duas regiões disjuntas, $\mathbf{I} = \{\mathbf{i}_1, \dots, \mathbf{i}_k\}$ e $\mathbf{J} = \{\mathbf{j}_1, \dots, \mathbf{j}_s\}$, introduzimos o madograma generalizado, que passamos a definir.

Definição 3.1 *Seja \mathbf{X} um campo aleatório max-estável com f.d. marginal univariada F de Fréchet unitária e $\mathbf{I} = \{\mathbf{i}_1, \dots, \mathbf{i}_k\}$ e $\mathbf{J} = \{\mathbf{j}_1, \dots, \mathbf{j}_s\}$ duas regiões disjuntas de \mathbb{R}^2 . O madograma generalizado define-se por*

$$\nu^{\alpha, \beta}(\mathbf{I}, \mathbf{J}) = \frac{1}{2} E \left| F^\alpha(M(\mathbf{I})) - F^\beta(M(\mathbf{J})) \right|, \quad \alpha > 0, \beta > 0, \quad (3.1)$$

onde $M(\mathbf{I}) = \bigvee_{p=1}^k X_{\mathbf{i}_p}$ e $M(\mathbf{J}) = \bigvee_{q=1}^s X_{\mathbf{j}_q}$.

Tomando $\beta = 1 - \alpha$, $\alpha \in (0, 1)$, e $k = s = 1$ em (3.1), obtém-se a definição do λ -madograma de Naveau *et al.* (2009) apresentada em (1.20).

Atendendo a que a f.d. F é Fréchet unitária o madograma generalizado pode ser reescrito da seguinte forma

$$\begin{aligned} \nu^{\alpha, \beta}(\mathbf{I}, \mathbf{J}) &= \frac{1}{2} E \left| \bigvee_{p=1}^k F^\alpha(X_{\mathbf{i}_p}) - \bigvee_{q=1}^s F^\beta(X_{\mathbf{j}_q}) \right| \\ &= \frac{1}{2} E \left| \bigvee_{p=1}^k F\left(\frac{X_{\mathbf{i}_p}}{\alpha}\right) - \bigvee_{q=1}^s F\left(\frac{X_{\mathbf{j}_q}}{\beta}\right) \right|. \end{aligned} \quad (3.2)$$

Esta expressão do madograma generalizado motivou o estimador que apresentaremos na Secção 3.3.

As constantes α e β representam as ponderações atribuídas às duas regiões de localizações. Uma escolha sensata, para quando não existe nenhuma localização a privilegiar em cada uma das regiões $\mathbf{I} = \{\mathbf{i}_1, \dots, \mathbf{i}_k\}$ e $\mathbf{J} = \{\mathbf{j}_1, \dots, \mathbf{j}_s\}$ e que generaliza a abordagem bivariada, é a seguinte

$$\alpha = \frac{\lambda}{k} \quad \text{e} \quad \beta = \frac{1 - \lambda}{s}, \quad \lambda \in (0, 1).$$

Na proposição que se segue mostra-se que $\nu^{\alpha, \beta}(\mathbf{I}, \mathbf{J})$ fornece informação sobre a dependência entre as regiões \mathbf{I} e \mathbf{J} através da função expoente associada à distribuição MEV G . Este resultado generaliza a Proposição 1. de Naveau *et al.* (2009).

Proposição 3.1 *Seja \mathbf{X} um campo aleatório max-estável com f.d. marginal univariada F de Fréchet unitária, $\mathbf{I} = \{\mathbf{i}_1, \dots, \mathbf{i}_k\}$ e $\mathbf{J} = \{\mathbf{j}_1, \dots, \mathbf{j}_s\}$ regiões disjuntas de localizações em \mathbb{R}^2 e $\alpha, \beta \in \mathbb{R}^+$. Então*

$$\nu^{\alpha, \beta}(\mathbf{I}, \mathbf{J}) = \frac{V_{\mathbf{I}, \mathbf{J}}(\alpha, \dots, \alpha, \beta, \dots, \beta)}{1 + V_{\mathbf{I}, \mathbf{J}}(\alpha, \dots, \alpha, \beta, \dots, \beta)} - c(\alpha, \beta) \quad (3.3)$$

com

$$c(\alpha, \beta) = \frac{1}{2} \left(\frac{V_{\mathbf{I}}(1, \dots, 1)}{\alpha + V_{\mathbf{I}}(1, \dots, 1)} + \frac{V_{\mathbf{J}}(1, \dots, 1)}{\beta + V_{\mathbf{J}}(1, \dots, 1)} \right),$$

onde, para $x_1, \dots, x_{k+s} \in \mathbb{R}^+$,

$$V_{\mathbf{I}, \mathbf{J}}(x_1, \dots, x_k, x_{k+1}, \dots, x_{k+s}) = -\ln G_{\mathbf{I}, \mathbf{J}}(x_1, \dots, x_k, x_{k+1}, \dots, x_{k+s})$$

e

$$G_{\mathbf{I}, \mathbf{J}}(x_1, \dots, x_{k+s}) = P \left(\bigcap_{p=1}^k \{X_{\mathbf{i}_p} \leq x_p\}, \bigcap_{q=1}^s \{X_{\mathbf{j}_q} \leq x_{k+q}\} \right).$$

Demonstração: O resultado obtém-se aplicando a relação

$$|a - b| = 2(a \vee b) - (a + b),$$

na definição de $\nu^{\alpha, \beta}(\mathbf{I}, \mathbf{J})$ e tendo em conta que

$$E \left(F^\alpha (M(\mathbf{I})) \vee F^\beta (M(\mathbf{J})) \right) = E \left(F \left(\frac{M(\mathbf{I})}{\alpha} \vee \frac{M(\mathbf{J})}{\beta} \right) \right).$$

De facto, atendendo a que a distribuição de $\frac{M(\mathbf{I})}{\alpha} \vee \frac{M(\mathbf{J})}{\beta}$ é dada por

$$\begin{aligned} P \left(\frac{M(\mathbf{I})}{\alpha} \vee \frac{M(\mathbf{J})}{\beta} \leq u \right) &= P (M(\mathbf{I}) \leq \alpha u, M(\mathbf{J}) \leq \beta u) \\ &= G_{\mathbf{I}, \mathbf{J}}(\alpha u, \dots, \alpha u, \beta u, \dots, \beta u) \\ &= \exp \left\{ -u^{-1} V_{\mathbf{I}, \mathbf{J}}(\alpha, \dots, \alpha, \beta, \dots, \beta) \right\}, \quad u > 0, \end{aligned}$$

obtém-se

$$\begin{aligned} &E \left(F \left(\frac{M(\mathbf{I})}{\alpha} \vee \frac{M(\mathbf{J})}{\beta} \right) \right) \\ &= \int_0^{+\infty} F(u) \exp \left\{ -V_{\mathbf{I}, \mathbf{J}}(\alpha u, \dots, \alpha u, \beta u, \dots, \beta u) \right\} \frac{d}{du} \left(-V_{\mathbf{I}, \mathbf{J}}(\alpha u, \dots, \alpha u, \beta u, \dots, \beta u) \right) du \end{aligned}$$

$$\begin{aligned}
 &= \int_0^{+\infty} \exp \left\{ -u^{-1} - u^{-1} V_{\mathbf{I}, \mathbf{J}}(\alpha, \dots, \alpha, \beta, \dots, \beta) \right\} u^{-2} V_{\mathbf{I}, \mathbf{J}}(\alpha, \dots, \alpha, \beta, \dots, \beta) du \\
 &= \frac{V_{\mathbf{I}, \mathbf{J}}(\alpha, \dots, \alpha, \beta, \dots, \beta)}{1 + V_{\mathbf{I}, \mathbf{J}}(\alpha, \dots, \alpha, \beta, \dots, \beta)} \times \\
 &\times \int_0^{+\infty} \exp \left\{ -u^{-1} (1 + V_{\mathbf{I}, \mathbf{J}}(\alpha, \dots, \alpha, \beta, \dots, \beta)) \right\} u^{-2} (1 + V_{\mathbf{I}, \mathbf{J}}(\alpha, \dots, \alpha, \beta, \dots, \beta)) du \\
 &= \frac{V_{\mathbf{I}, \mathbf{J}}(\alpha, \dots, \alpha, \beta, \dots, \beta)}{1 + V_{\mathbf{I}, \mathbf{J}}(\alpha, \dots, \alpha, \beta, \dots, \beta)}
 \end{aligned}$$

De forma análoga, deduz-se que

$$E(F^\alpha(M(\mathbf{I}))) = \frac{V_{\mathbf{I}}(\alpha, \dots, \alpha)}{1 + V_{\mathbf{I}}(\alpha, \dots, \alpha)} = \frac{V_{\mathbf{I}}(1, \dots, 1)}{\alpha + V_{\mathbf{I}}(1, \dots, 1)}$$

e

$$E(F^\beta(M(\mathbf{J}))) = \frac{V_{\mathbf{J}}(\beta, \dots, \beta)}{1 + V_{\mathbf{J}}(\beta, \dots, \beta)} = \frac{V_{\mathbf{J}}(1, \dots, 1)}{\beta + V_{\mathbf{J}}(1, \dots, 1)}.$$

□

A relação estabelecida na proposição anterior evidencia que $\nu^{\alpha, \beta}(\mathbf{I}, \mathbf{J})$ tem em consideração a dependência dentro de cada uma das região \mathbf{I} e \mathbf{J} , por via dos coeficientes extremais dos vetores $(X_{\mathbf{i}_1}, \dots, X_{\mathbf{i}_k})$ e $(X_{\mathbf{j}_1}, \dots, X_{\mathbf{j}_s})$, e entre as regiões \mathbf{I} e \mathbf{J} através da função expoente $V_{\mathbf{I}, \mathbf{J}}(\alpha, \dots, \alpha, \beta, \dots, \beta)$.

Na proposição seguinte apresentam-se algumas propriedades do madograma generalizado.

Proposição 3.2 *Sejam $\mathbf{I} = \{\mathbf{i}_1, \dots, \mathbf{i}_k\}$ e $\mathbf{J} = \{\mathbf{j}_1, \dots, \mathbf{j}_s\}$ regiões disjuntas de \mathbb{R}^2 . Para cada $\alpha, \beta \in \mathbb{R}^+$, tem-se:*

1. $0 \leq \nu^{\alpha, \beta}(\mathbf{I}, \mathbf{J}) \leq \frac{1}{2}$;
2. $\nu^{\alpha, \alpha}(\mathbf{I}, \mathbf{J}) = \frac{\varepsilon(\mathbf{I} \cup \mathbf{J})}{\alpha + \varepsilon(\mathbf{I} \cup \mathbf{J})} - \frac{1}{2} \left(\frac{\varepsilon(\mathbf{I})}{\alpha + \varepsilon(\mathbf{I})} + \frac{\varepsilon(\mathbf{J})}{\alpha + \varepsilon(\mathbf{J})} \right)$.

Demonstração:

1. O intervalo de variação obtém-se diretamente da definição de madograma generalizado.
2. O resultado é imediato a partir da relação $\varepsilon(\mathbf{I} \cup \mathbf{J}) = V_{\mathbf{I}, \mathbf{J}}(1, \dots, 1)$.

□

Aproveitamos o resultado apresentado no ponto **2.** da Proposição 3.2 para pôr em destaque algumas questões.

Observação 3.1 *Se as variáveis $M(\mathbf{I})$ e $M(\mathbf{J})$ forem independentes tem-se*

$$\nu^{\alpha,\alpha}(\mathbf{I}, \mathbf{J}) = \frac{\varepsilon(\mathbf{I}) + \varepsilon(\mathbf{J})}{\alpha + \varepsilon(\mathbf{I}) + \varepsilon(\mathbf{J})} - \frac{1}{2} \left(\frac{\varepsilon(\mathbf{I})}{\alpha + \varepsilon(\mathbf{I})} + \frac{\varepsilon(\mathbf{J})}{\alpha + \varepsilon(\mathbf{J})} \right),$$

e no caso de serem totalmente dependentes obtém-se

$$\nu^{\alpha,\alpha}(\mathbf{I}, \mathbf{J}) = \frac{\varepsilon(\mathbf{I}) \vee \varepsilon(\mathbf{J})}{\alpha + (\varepsilon(\mathbf{I}) \vee \varepsilon(\mathbf{J}))} - \frac{1}{2} \left(\frac{\varepsilon(\mathbf{I})}{\alpha + \varepsilon(\mathbf{I})} + \frac{\varepsilon(\mathbf{J})}{\alpha + \varepsilon(\mathbf{J})} \right).$$

Observação 3.2 *O coeficiente de dependência $\nu^{\alpha,\alpha}(\mathbf{I}, \mathbf{J})$ relaciona-se com os coeficientes de dependência (1.13) e (1.14) da seguinte forma:*

$$\nu^{\alpha,\alpha}(\mathbf{I}, \mathbf{J}) = \frac{\varepsilon(\mathbf{I})\varepsilon_1(\mathbf{I}, \mathbf{J})}{\alpha + \varepsilon(\mathbf{I})\varepsilon_1(\mathbf{I}, \mathbf{J})} - c(\alpha, \alpha) = \frac{(\varepsilon(\mathbf{J}) + \varepsilon(\mathbf{I}))\varepsilon_2(\mathbf{I}, \mathbf{J})}{\alpha + (\varepsilon(\mathbf{J}) + \varepsilon(\mathbf{I}))\varepsilon_2(\mathbf{I}, \mathbf{J})} - c(\alpha, \alpha), \quad \alpha > 0.$$

3.3 Estimação

A Proposição 3.1 relaciona o madograma generalizado com a função expoente. Denotando por $G_{\mathbf{A}}$, com $\mathbf{A} = \{\mathbf{i}_1, \dots, \mathbf{i}_k\}$, a distribuição MEV do vetor $(X_{\mathbf{i}_1}, \dots, X_{\mathbf{i}_k})$ e por $D_{G_{\mathbf{A}}}$ a sua função de dependência, a função de dependência na cauda associada à região \mathbf{A} define-se por

$$l_{\mathbf{A}}(x_1, \dots, x_k) = \lim_{t \downarrow 0} t^{-1} (1 - D_{G_{\mathbf{A}}}(1 - tx_1, \dots, 1 - tx_k)), \quad (3.4)$$

com $x_i \in \mathbb{R}^+$, $i \in \{1, \dots, k\}$, e relaciona-se com a função expoente do seguinte modo

$$l_{\mathbf{A}}(1/x_1, \dots, 1/x_k) = V_{\mathbf{A}}(x_1, \dots, x_k), \quad x_i \in \mathbb{R}^+, i \in \{1, \dots, k\}. \quad (3.5)$$

Existem várias referências na literatura relativamente à estimação da função l , de que são exemplo Einmahl (2012), Krajina (2010) e Beirlant (2004). Qualquer um dos estimadores aí referenciados pode ser aplicado na estimação do madograma generalizado.

Nesta secção propomos um estimador não paramétrico para $\nu^{\alpha,\beta}$. Usualmente, os estimadores não paramétricos envolvem a escolha de um valor k que corresponde ao número de observações de topo que deverá ser utilizado na estimação de um quantil elevado de F . Neste caso, temos o problema usual da variância e do viés: um valor de k demasiado baixo leva-nos a um estimador com elevada variância e um valor

demasiado elevado introduz viés. Alguns métodos que auxiliam na escolha de k , de forma a conseguir-se um bom compromisso entre variância e viés, são apresentados em Einmahl *et al.* (2006).

O estimador que propomos para o madograma generalizado não envolve a escolha de k . Provaremos a sua consistência e normalidade assintótica tanto no caso em que a f.d. das variáveis do campo \mathbf{X} é conhecida como no caso em que é desconhecida.

Sejam

$$(X_{\mathbf{i}_1}^{(t)}, \dots, X_{\mathbf{i}_k}^{(t)}) \text{ e } (X_{\mathbf{j}_1}^{(t)}, \dots, X_{\mathbf{j}_s}^{(t)}), \quad t = 1, \dots, T,$$

réplicas independentes dos vetores $(X_{\mathbf{i}_1}, \dots, X_{\mathbf{i}_k})$ e $(X_{\mathbf{j}_1}, \dots, X_{\mathbf{j}_s})$, respetivamente. Então,

$$\left\{ M_t(\mathbf{I}) = \bigvee_{p=1}^k X_{\mathbf{i}_p}^{(t)}, \quad t = 1, \dots, T \right\} \text{ e } \left\{ M_t(\mathbf{J}) = \bigvee_{q=1}^s X_{\mathbf{j}_q}^{(t)}, \quad t = 1, \dots, T \right\}$$

são amostras aleatórias de $M(\mathbf{I})$ e $M(\mathbf{J})$, respetivamente.

1º caso: Função de distribuição conhecida.

Supondo conhecida a f.d. de $X_{\mathbf{i}_p}$, $F_{\mathbf{i}_p}$, $\mathbf{i}_p \in \mathbb{R}^2$, a qual poderá ser transformada em Fréchet unitária considerando

$$-\frac{1}{\ln F_{\mathbf{i}_p}(X_{\mathbf{i}_p})} \quad (3.6)$$

propomos o seguinte estimador para o madograma generalizado

$$\widehat{\nu}^{\alpha, \beta}(\mathbf{I}, \mathbf{J}) = \frac{1}{2T} \sum_{i=1}^T |F^\alpha(M_i(\mathbf{I})) - F^\beta(M_i(\mathbf{J}))|, \quad \alpha > 0, \beta > 0. \quad (3.7)$$

A consistência deste estimador e a sua normalidade assintótica apresentam-se na proposição seguinte.

Proposição 3.3 *Considerando o estimador definido em (3.7), para o madograma generalizado, tem-se*

$$\frac{\sqrt{T}(\widehat{\nu}^{\alpha, \beta}(\mathbf{I}, \mathbf{J}) - \nu^{\alpha, \beta}(\mathbf{I}, \mathbf{J}))}{\sigma} \rightarrow N(0, 1),$$

onde $\sigma^2 = \frac{1}{2}\gamma_F^{\alpha, \beta}(\mathbf{I}, \mathbf{J}) - (\nu^{\alpha, \beta}(\mathbf{I}, \mathbf{J}))^2$ e $\gamma_F^{\alpha, \beta}(\mathbf{I}, \mathbf{J}) = \frac{1}{2}E[(F^\alpha(M(\mathbf{I})) - F^\beta(M(\mathbf{J})))^2]$. Além disso, $\widehat{\nu}^{\alpha, \beta}(\mathbf{I}, \mathbf{J})$ converge quase certamente para $\nu^{\alpha, \beta}(\mathbf{I}, \mathbf{J})$.

Demonstração: Sejam Y_1, \dots, Y_T cópias independentes de

$$Y = \frac{1}{2} |F^\alpha(M(\mathbf{I})) - F^\beta(M(\mathbf{J}))|.$$

Então

$$\frac{\sqrt{T}(\bar{Y} - \mu_Y)}{\sigma_Y} \rightarrow N(0, 1),$$

onde

$$\mu_Y = \frac{1}{2} E |F^\alpha(M(\mathbf{I})) - F^\beta(M(\mathbf{J}))| = \nu^{\alpha,\beta}(\mathbf{I}, \mathbf{J}) \quad \text{e} \quad \sigma_Y^2 = \frac{1}{2} \gamma_F^{\alpha,\beta}(\mathbf{I}, \mathbf{J}) - (\nu^{\alpha,\beta}(\mathbf{I}, \mathbf{J}))^2.$$

A consistência forte de $\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$ obtém-se dado que a média amostral converge quase certamente para o valor médio. \square

2º caso: Função de distribuição desconhecida.

O desconhecimento, na prática, da f.d. marginal univariada obriga à sua estimação, para o que recorremos à correspondente função empírica.

Substituindo, para cada $\mathbf{i}_p \in \mathbb{R}^2$, a desconhecida marginal $F_{\mathbf{i}_p}$ pela homóloga empírica e considerando a transformação das variáveis $X_{\mathbf{i}_p}, \mathbf{i}_p \in \mathbb{R}^2$, em variáveis

$$\widehat{U}_{\mathbf{i}_p}^{(t)} = -\frac{1}{\ln(\widehat{F}_{\mathbf{i}_p}(X_{\mathbf{i}_p}^{(t)}))},$$

com distribuição aproximada de Fréchet unitária, obtemos o estimador

$$\begin{aligned} \widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J}) &= \frac{1}{2T} \sum_{t=1}^T \left| F^\alpha \left(\bigvee_{p=1}^k \widehat{U}_{\mathbf{i}_p}^{(t)} \right) - F^\beta \left(\bigvee_{q=1}^s \widehat{U}_{\mathbf{j}_q}^{(t)} \right) \right| \\ &= \frac{1}{2T} \sum_{t=1}^T \left| \bigvee_{p=1}^k F \left(\frac{\widehat{U}_{\mathbf{i}_p}^{(t)}}{\alpha} \right) - \bigvee_{q=1}^s F \left(\frac{\widehat{U}_{\mathbf{j}_q}^{(t)}}{\beta} \right) \right| \\ &= \frac{1}{2T} \sum_{t=1}^T \left| \bigvee_{p=1}^k \widehat{F}_{\mathbf{i}_p}^\alpha \left(X_{\mathbf{i}_p}^{(t)} \right) - \bigvee_{q=1}^s \widehat{F}_{\mathbf{j}_q}^\beta \left(X_{\mathbf{j}_q}^{(t)} \right) \right|, \end{aligned} \quad (3.8)$$

com $\alpha > 0, \beta > 0$ e

$$\widehat{F}_{\mathbf{i}}(u) = \frac{1}{n+1} \sum_{j=1}^n \mathbb{1}_{\{X_{\mathbf{i}}^{(j)} \leq u\}}, \quad (3.9)$$

a função empírica modificada, onde se considera $n+1$ em vez de n , para corrigir a precisão da estimativa como é sugerido em Beirlant *et al.* (2004).

Também este estimador é fortemente consistente, como se mostra na proposição

seguinte.

Proposição 3.4 *O estimador $\widehat{\mathcal{V}}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$, definido em (3.8), é fortemente consistente.*

Demonstração: O estimador $\widehat{\mathcal{V}}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$ é fortemente consistente dado que

$$\begin{aligned} & \left| \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \left| \bigvee_{p=1}^k \widehat{F}_{i_p}^{\alpha}(X_{i_p}^{(t)}) - \bigvee_{q=1}^s \widehat{F}_{j_q}^{\beta}(X_{j_q}^{(t)}) \right| - \frac{1}{2} E \left| \bigvee_{q=1}^s F^{\alpha}(X_{j_q}^{(t)}) - \bigvee_{q=1}^s F^{\beta}(X_{j_q}^{(t)}) \right| \right| \\ & \leq \left| \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \left| \bigvee_{p=1}^k \widehat{F}_{i_p}^{\alpha}(X_{i_p}^{(t)}) - \bigvee_{q=1}^s \widehat{F}_{j_q}^{\beta}(X_{j_q}^{(t)}) \right| - \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \left| \bigvee_{p=1}^k F^{\alpha}(X_{i_p}^{(t)}) - \bigvee_{q=1}^s F^{\beta}(X_{j_q}^{(t)}) \right| \right| \\ & \quad + \left| \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \left| \bigvee_{p=1}^k F^{\alpha}(X_{i_p}^{(t)}) - \bigvee_{q=1}^s F^{\beta}(X_{j_q}^{(t)}) \right| - \frac{1}{2} E \left| \bigvee_{p=1}^k F^{\alpha}(X_{i_p}^{(t)}) - \bigvee_{q=1}^s F^{\beta}(X_{j_q}^{(t)}) \right| \right|. \end{aligned}$$

De facto, aplicando a Lei Forte dos Grandes Números prova-se que o segundo termo converge quase certamente para zero. Relativamente ao primeiro termo tem-se

$$\begin{aligned} & \left| \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \left| \bigvee_{p=1}^k \widehat{F}_{i_p}^{\alpha}(X_{i_p}^{(t)}) - \bigvee_{q=1}^s \widehat{F}_{j_q}^{\beta}(X_{j_q}^{(t)}) \right| - \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \left| \bigvee_{p=1}^k F^{\alpha}(X_{i_p}^{(t)}) - \bigvee_{q=1}^s F^{\beta}(X_{j_q}^{(t)}) \right| \right| \\ & \leq \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \left| \left| \bigvee_{p=1}^k \widehat{F}_{i_p}^{\alpha}(X_{i_p}^{(t)}) - \bigvee_{q=1}^s \widehat{F}_{j_q}^{\beta}(X_{j_q}^{(t)}) \right| - \left| \bigvee_{p=1}^k F^{\alpha}(X_{i_p}^{(t)}) - \bigvee_{q=1}^s F^{\beta}(X_{j_q}^{(t)}) \right| \right| \\ & \leq \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \left| \bigvee_{p=1}^k \widehat{F}_{i_p}^{\alpha}(X_{i_p}^{(t)}) - \bigvee_{q=1}^s \widehat{F}_{j_q}^{\beta}(X_{j_q}^{(t)}) - \bigvee_{p=1}^k F^{\alpha}(X_{i_p}^{(t)}) + \bigvee_{q=1}^s F^{\beta}(X_{j_q}^{(t)}) \right| \\ & \leq \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \left[\bigvee_{p=1}^k \left| \widehat{F}_{i_p}^{\alpha}(X_{i_p}^{(t)}) - F^{\alpha}(X_{i_p}^{(t)}) \right| + \bigvee_{q=1}^s \left| \widehat{F}_{j_q}^{\beta}(X_{j_q}^{(t)}) - F^{\beta}(X_{j_q}^{(t)}) \right| \right] \\ & \leq \frac{1}{2} \frac{1}{T} \sum_{t=1}^T \left[\sum_{p=1}^k \left| \widehat{F}_{i_p}^{\alpha}(X_{i_p}^{(t)}) - F^{\alpha}(X_{i_p}^{(t)}) \right| + \sum_{q=1}^s \left| \widehat{F}_{j_q}^{\beta}(X_{j_q}^{(t)}) - F^{\beta}(X_{j_q}^{(t)}) \right| \right], \end{aligned}$$

que converge quase certamente para zero de acordo com Gilat e Hill (1992) (demonstração do Teorema 1.1). \square

A normalidade assintótica do estimador dado em (3.8) obtém-se do teorema que se segue, apresentado em Fermanian *et al.* (2004), considerando a função $J : [0, 1]^{k+s} \rightarrow \mathbb{R}$ definida do seguinte modo

$$J(x_1, \dots, x_{k+s}) = \frac{1}{2} \left| \bigvee_{p=1}^k x_p^{\alpha} - \bigvee_{p=k+1}^{k+s} x_p^{\beta} \right|.$$

Teorema 3.1 *Seja $(X_{i_1}, \dots, X_{i_k}, X_{j_1}, \dots, X_{j_s})$ um vetor aleatório com f.d. H e com f. d.'s marginais contínuas $F_{i_1}, \dots, F_{i_k}, F_{j_1}, \dots, F_{j_s}$ e cuja função de dependência associada a H , D_H , tem derivadas parciais contínuas. Se $J : [0, 1]^{k+s} \rightarrow \mathbb{R}$ é uma função de variação limitada, contínua por valores superiores e com descontinuidades de 1ª espécie, então quando $T \rightarrow +\infty$, tem-se a seguinte convergência em distribuição em $l^\infty([0, 1]^{k+s})$*

$$\begin{aligned} & \frac{1}{\sqrt{T}} \sum_{t=1}^T \left\{ J(\widehat{F}_{i_1}(X_{i_1}^{(t)}), \dots, \widehat{F}_{i_k}(X_{i_k}^{(t)}), \widehat{F}_{j_1}(X_{j_1}^{(t)}), \dots, \widehat{F}_{j_s}(X_{j_s}^{(t)})) \right. \\ & \quad \left. - E(J(F_{i_1}(X_{i_1}^{(t)}), \dots, F_{i_k}(X_{i_k}^{(t)}), F_{j_1}(X_{j_1}^{(t)}), \dots, F_{j_s}(X_{j_s}^{(t)}))) \right\} \\ & \rightarrow \int_{[0,1]^{k+s}} \mathbb{G}(u_1, \dots, u_{k+s}) dJ(u_1, \dots, u_{k+s}), \end{aligned}$$

onde o processo limite e \mathbb{G} têm distribuição Gaussiana com média zero.

Na próxima secção avaliamos o desempenho do estimador dado em (3.8), através de um estudo de simulação.

3.4 Estudo de simulação: alguns exemplos

A classe dos campos max-estáveis M4, apresentada na Secção 2.6, será utilizada para avaliar o desempenho do estimador (3.8), uma vez que definida uma estrutura de dependência é possível determinar o valor exato do madograma generalizado (3.1) e compará-lo com a estimativa obtida a partir de (3.8).

Seja $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{R}^2}$, o campo M4 definido na Secção 2.6. Atendendo a (2.22) o madograma generalizado do campo M4 é dado por

$$\nu^{\alpha, \beta}(\mathbf{I}, \mathbf{J}) = \frac{\sum_{l=1}^{+\infty} \sum_{m=-\infty}^{+\infty} \left(\bigvee_{p=1}^k \alpha^{-1} a_{lm\mathbf{i}_p} \bigvee_{p=1}^s \beta^{-1} a_{lm\mathbf{j}_p} \right)}{1 + \sum_{l=1}^{+\infty} \sum_{m=-\infty}^{+\infty} \left(\bigvee_{p=1}^k \alpha^{-1} a_{lm\mathbf{i}_p} \bigvee_{p=1}^s \beta^{-1} a_{lm\mathbf{j}_p} \right)} - c(\alpha, \beta)$$

com

$$c(\alpha, \beta) = \frac{1}{2} \left(\frac{\sum_{l=1}^{+\infty} \sum_{m=-\infty}^{+\infty} \bigvee_{p=1}^k a_{lm\mathbf{i}_p}}{\alpha + \sum_{l=1}^{+\infty} \sum_{m=-\infty}^{+\infty} \bigvee_{p=1}^k a_{lm\mathbf{i}_p}} + \frac{\sum_{l=1}^{+\infty} \sum_{m=-\infty}^{+\infty} \bigvee_{p=1}^s a_{lm\mathbf{j}_p}}{\beta + \sum_{l=1}^{+\infty} \sum_{m=-\infty}^{+\infty} \bigvee_{p=1}^s a_{lm\mathbf{j}_p}} \right).$$

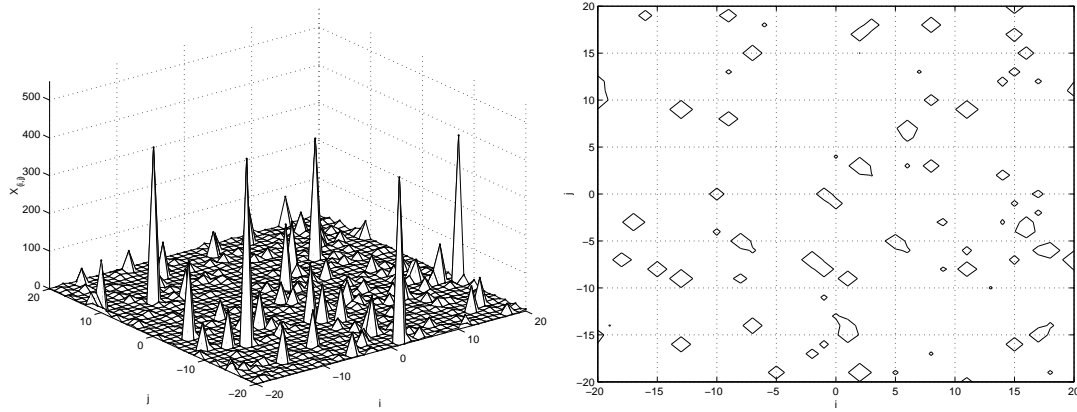


Figura 3.1: Simulação do campo aleatório M4 definido no Exemplo 3.1 (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 18.5626$ (direita).

Nos dois exemplos que se seguem obtemos a expressão do madograma generalizado de um campo M4, $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{Z}^2}$, com 1 e 2 padrões de assinaturas, para as regiões de localizações $\mathbf{I} = \{(2, 1), (2, 2)\}$ e $\mathbf{J} = \{(2, 3), (3, 3)\}$.

Exemplo 3.1 Para cada localização $\mathbf{i} = (i_1, i_2) \in \mathbb{Z}^2$ consideramos as constantes definidas por

$$\begin{cases} a_{11\mathbf{i}} = a_{12\mathbf{i}} = \frac{1}{2} & \text{se } i_1 \text{ e } i_2 \text{ são pares} \\ a_{11\mathbf{i}} = \frac{1}{4} = 1 - a_{12\mathbf{i}} & \text{caso contrário} \end{cases}.$$

Os valores de $(a_{11\mathbf{i}}, a_{12\mathbf{i}})$ determinam o padrão móvel ou o padrão de assinaturas do campo aleatório (Figura 3.1).

Para as regiões de localizações disjuntas \mathbf{I} e \mathbf{J} o madograma generalizado é dado por

$$\nu^{\alpha, \beta}(\mathbf{I}, \mathbf{J}) = \frac{\frac{1}{4}(2\alpha^{-1} \vee \beta^{-1}) + \frac{3}{4}(\alpha^{-1} \vee \beta^{-1})}{1 + \frac{1}{4}(2\alpha^{-1} \vee \beta^{-1}) + \frac{3}{4}(\alpha^{-1} \vee \beta^{-1})} - \frac{1}{2} \left(\frac{\frac{5}{4}}{\alpha + \frac{5}{4}} + \frac{1}{\beta + 1} \right), \quad \alpha > 0, \beta > 0.$$

Exemplo 3.2 Neste exemplo consideramos dois padrões de assinaturas, $L = 2$. Para cada localização, $\mathbf{i} = (i_1, i_2) \in \mathbb{Z}^2$, as constantes são definidas do seguinte modo

$$\begin{cases} a_{11\mathbf{i}} = a_{12\mathbf{i}} = a_{13\mathbf{i}} = \frac{1}{12}, a_{21\mathbf{i}} = a_{22\mathbf{i}} = a_{23\mathbf{i}} = \frac{1}{4} & \text{se } i_1 \text{ e } i_2 \text{ são ímpares} \\ a_{11\mathbf{i}} = \frac{1}{18}, a_{12\mathbf{i}} = \frac{1}{9}, a_{13\mathbf{i}} = \frac{1}{6}, a_{21\mathbf{i}} = a_{22\mathbf{i}} = a_{23\mathbf{i}} = \frac{2}{9} & \text{caso contrário} \end{cases}.$$

Assim, $(a_{11\mathbf{i}}, a_{12\mathbf{i}}, a_{13\mathbf{i}})$ e $(a_{21\mathbf{i}}, a_{22\mathbf{i}}, a_{23\mathbf{i}})$ definem os dois padrões de assinaturas do campo aleatório (Figura 3.2).

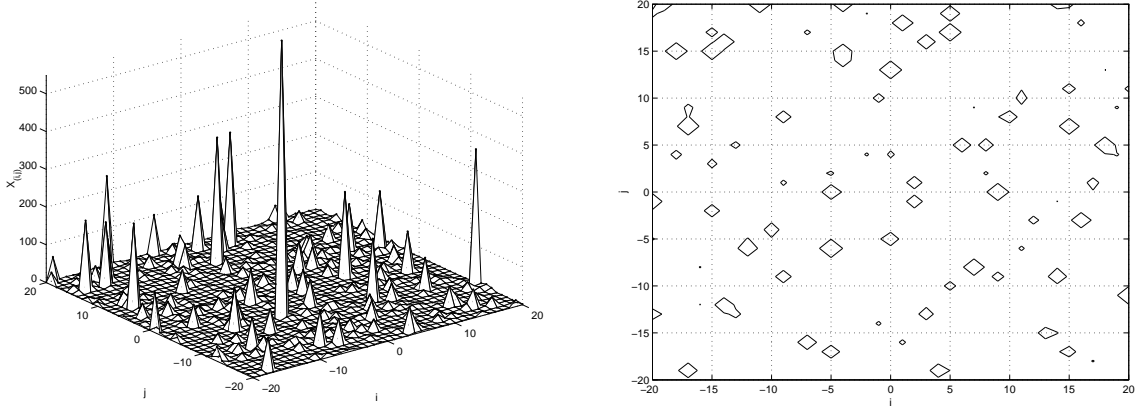


Figura 3.2: Simulação do campo aleatório M4 definido no Exemplo 3.2 (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 20.2763$ (direita).

Para as regiões **I** e **J** obtém-se

$$\nu^{\alpha,\beta}(\mathbf{I}, \mathbf{J}) = \frac{\left(\frac{\alpha^{-1}}{18} \sqrt{\frac{\beta^{-1}}{12}}\right) + \frac{(\alpha^{-1}\sqrt{\beta^{-1}})}{9} + \frac{(\alpha^{-1}\sqrt{\beta^{-1}})}{6} + \left(\frac{2\alpha^{-1}}{3} \sqrt{\frac{3\beta^{-1}}{4}}\right)}{1 + \left(\frac{\alpha^{-1}}{18} \sqrt{\frac{\beta^{-1}}{12}}\right) + \frac{(\alpha^{-1}\sqrt{\beta^{-1}})}{9} + \frac{(\alpha^{-1}\sqrt{\beta^{-1}})}{6} + \left(\frac{2\alpha^{-1}}{3} \sqrt{\frac{3\beta^{-1}}{4}}\right)} - \frac{1}{2} \left(\frac{1}{\alpha + 1} + \frac{\frac{10}{9}}{\beta + \frac{10}{9}}\right), \quad \alpha > 0, \beta > 0.$$

Os exemplos apresentados são utilizados nos estudos de simulação, que se apresentam de seguida, no sentido de avaliar o desempenho do estimador (3.8).

As Figuras 3.3 e 3.5 ilustram as simulações obtidas com a geração de 50 réplicas de dimensão 100 ($T = 100$) dos campos M4 i.i.d. dos modelos apresentados nos Exemplos 3.1 e 3.2, onde se considerou α e β com valores em $\{0.2, 0.4, 0.6, 0.8, 1, 2, 3, \dots, 20\}$. Com vista a uma melhor visualização do comportamento do estimador representam-se, nas Figuras 3.4 e 3.6, os valores exatos e a média dos valores estimados do madograma generalizado para alguns valores de α e β .

Observando os valores do erro quadrático médio (Figuras 3.3 e 3.5) podemos concluir que as estimativas fornecidas pelo estimador $\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$ estão muito próximas dos valores exatos do madograma generalizado. Esta conclusão é reforçada pelas Figuras 3.4 e 3.6, onde os desvios entre os valores exatos e a média dos estimados são imperceptíveis. Refira-se, contudo, que as piores estimativas no modelo com um padrão de assinaturas (Figura 3.3) ocorreram quando $\beta > 10$ e $\alpha < \beta$, o que corresponde a uma maior ponderação das localizações da região **J**. No caso do modelo com dois padrões de assinaturas (Figura 3.5) o maior erro quadrático médio observa-se quando $\alpha = \beta$.

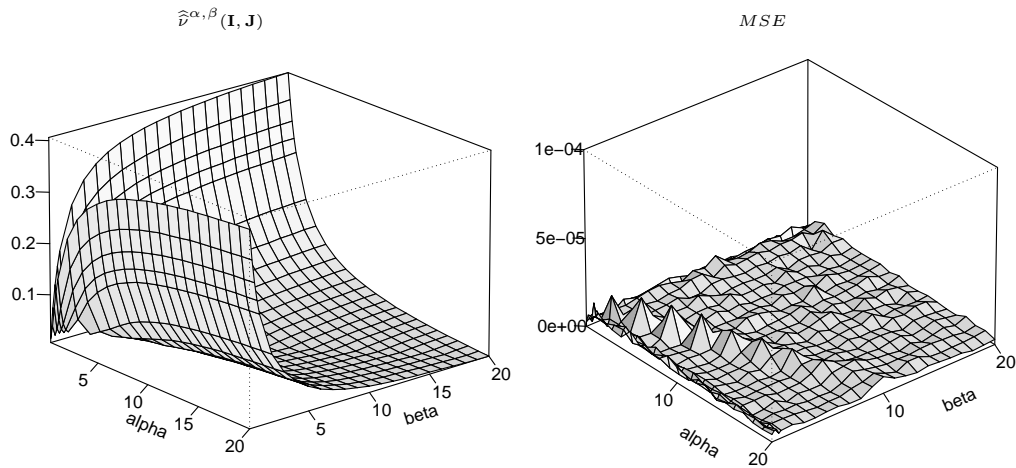


Figura 3.3: Média dos valores estimados do madograma generalizado ($\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$) e erro quadrático médio (MSE) para o Exemplo 3.1, com $\alpha, \beta \in \{0.2, 0.4, 0.6, 0.8, 1, 2, 3, \dots, 20\}$.

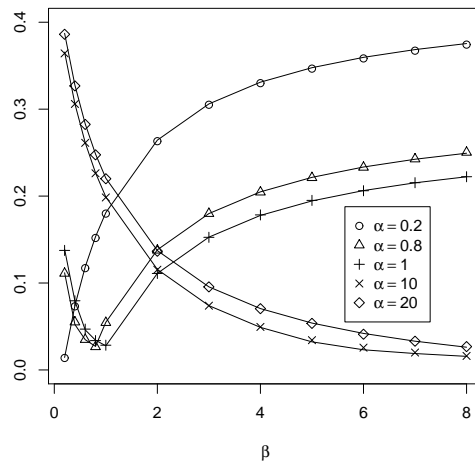


Figura 3.4: Valores exatos do madograma generalizado $\nu^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$ (linha contínua) e média dos valores estimados $\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$ (pontos “o, Δ, ◇, ×, +”) para o Exemplo 3.1, com $\alpha \in \{0.2, 0.8, 1, 10, 20\}$ e $\beta \in \{0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5, 6, 7, 8\}$.

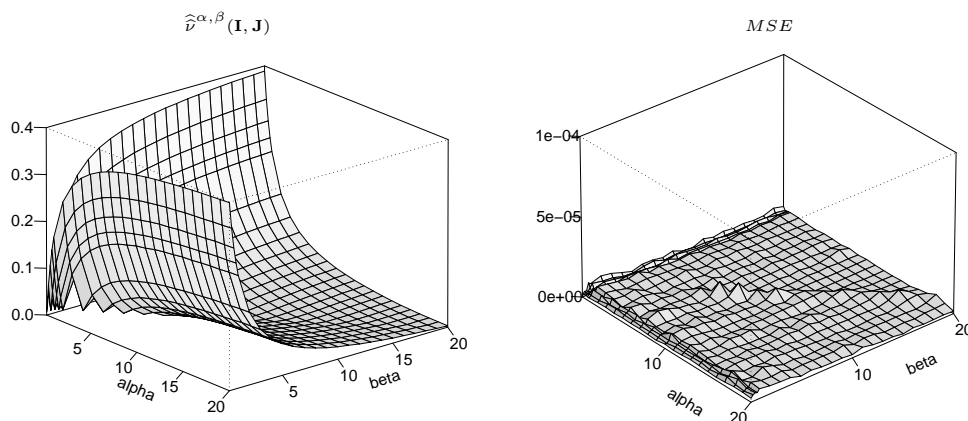


Figura 3.5: Média dos valores estimados do madograma generalizado ($\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$) e erro quadrático médio (MSE) para o Exemplo 3.2, com $\alpha, \beta \in \{0.2, 0.4, 0.6, 0.8, 1, 2, 3, \dots, 20\}$.

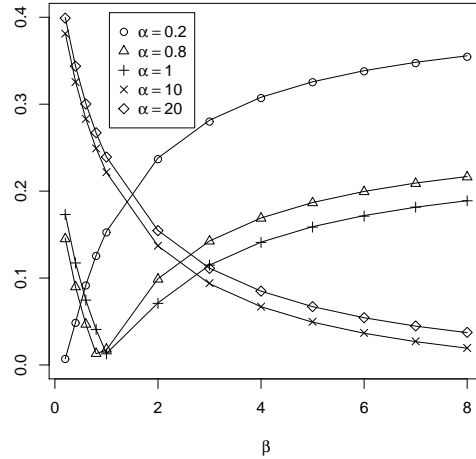


Figura 3.6: Valores exatos do madograma generalizado $\nu^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$ (linha contínua) e média dos valores estimados $\widehat{\nu}^{\alpha,\beta}(\mathbf{I}, \mathbf{J})$ (pontos “o, Δ, ◇, ×, +”) para o Exemplo 3.2, com $\alpha \in \{0.2, 0.8, 1, 10, 20\}$ e $\beta \in \{0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5, 6, 7, 8\}$.

3.5 Aplicação a dados de precipitação

Nesta secção, apresentam-se os resultados da aplicação do estimador proposto para o madograma generalizado, a dados de precipitação máxima anual de máximos diários em diferentes regiões topográficas de Portugal.

Em Portugal existem diferenças topográficas significativas entre o norte e o sul. O norte é caracterizado por ser muito montanhoso, enquanto no sul existem sobretudo planaltos e planícies. A Cordilheira Central, formada pelas serras de Sintra, Montejunto e Estrela, divide o sul e o norte e influencia o clima de Portugal, nomeadamente em termos de precipitação. Grande parte da precipitação em Portugal vem da direcção noroeste, sendo mais abundante no norte do que no sul, sendo um dos factores a barreira física à precipitação criada pela Cordilheira Central.

Para avaliar a dependência entre valores extremos de precipitação nesta zona montanhosa e nas áreas vizinhas, consideraram-se os valores máximos de precipitação anual de máximos diários registados ao longo de 32 anos (1944-1952, 1954-1965, 1967-1969, 1971, 1974-1975, 1977-1981), em 5 estações portuguesas, obtidos no SNIRH. Refira-se que a amostra considerada foi condicionada pelos registos disponíveis. A Figura 3.7 apresenta a localização das 5 estações consideradas. As estações “Lagoa Comprida” e “Fajão” estão localizadas a noroeste na Serra da Estrela, que pertence à Cordilheira Central e é a montanha mais alta de Portugal Continental. A norte da Cordilheira Central situa-se a estação “Caldas da Felgueira” e a sul as estações “Barragem Cabeço Monteiro”, “Penamacor”.

Atendendo a que os dados são máximos relativos a um longo período de tempo, assumimos que em cada localização são independentes ao longo dos anos. Supomos, ainda, que o campo aleatório é max-estável com distribuição marginal univariada

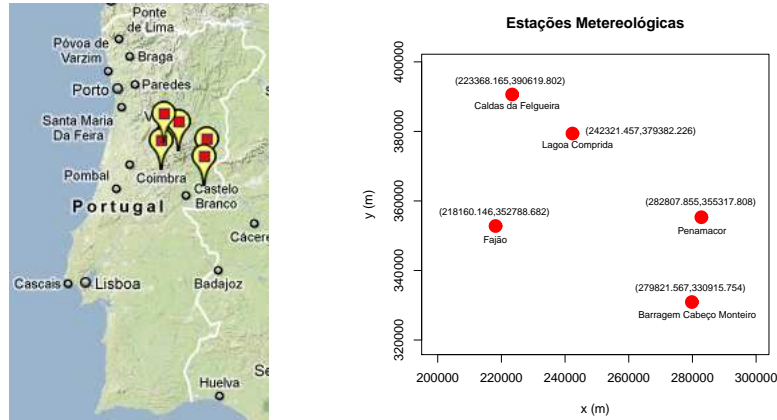


Figura 3.7: Localização das estações, obtida a partir do SNIRH (esquerda), e a sua representação em coordenadas Lambert (direita).

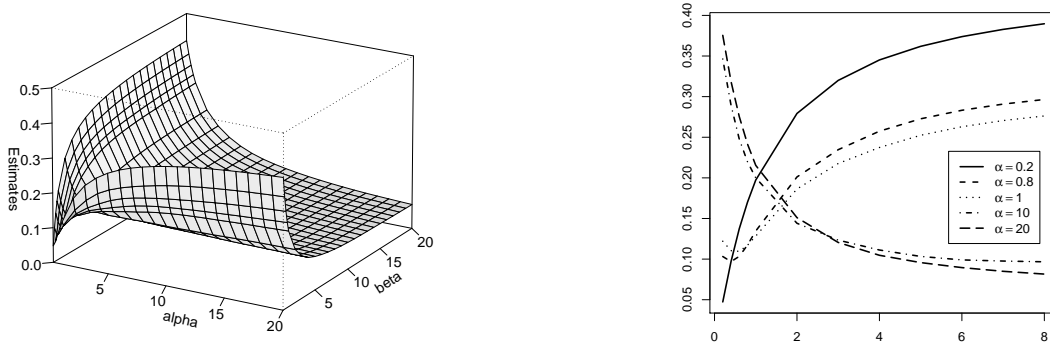


Figura 3.8: Estimativas do madograma generalizado $(\hat{\nu}^{\alpha, \beta}(\mathbf{I}, \mathbf{J}))$ obtidas para $\mathbf{I} = \{\text{L.Comprida, Fajão}\}$, $\mathbf{J} = \{\text{B.C.M.}\}$, com $\alpha, \beta \in \{0.2, 0.4, 0.6, 0.8, 1, 2, 3, \dots, 20\}$, à esquerda, e para $\alpha \in \{0.2, 0.8, 1, 10, 20\}$ e $\beta \in \{0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5, 6, 7, 8\}$, à direita.

desconhecida.

Nas Figuras 3.8, 3.9 e 3.10 representam-se os valores estimados do madograma generalizado, para diferentes valores de α e β , quando se considera a região montanhosa $\mathbf{I} = \{\text{Fajão, L.Comprida}\}$ e três outras regiões vizinhas, uma a norte ($\mathbf{J} = \{\text{C.Felgueira}\}$) e outra a sul ($\mathbf{J} = \{\text{Penamacor}\}$). Os resultados mostram que os valores mais baixos ocorrem entre a região montanhosa e “C. Felgueira” e os mais elevados entre a região montanhosa e “Penamacor”, sugerindo maior dependência no primeiro caso.

Dado que não temos motivos para ponderar de forma diferente as localizações em cada região, considerámos $\alpha = \lambda/k$ e $\beta = (1 - \lambda)/s$, $\lambda \in (0, 1)$, onde k e s são o número de localizações das regiões \mathbf{I} e \mathbf{J} , respetivamente. Nestas condições, aplicámos o estimador $\hat{\nu}^{\alpha, \beta}(\mathbf{I}, \mathbf{J})$ para quantificar a dependência entre as regiões definidas anteriormente e obtiveram-se os resultados apresentados na Figura 3.11.

Dado que valores baixos de $\nu^{\alpha, \beta}(\mathbf{I}, \mathbf{J})$ indicam dependência forte, os resultados re-

3.5. APLICAÇÃO A DADOS DE PRECIPITAÇÃO

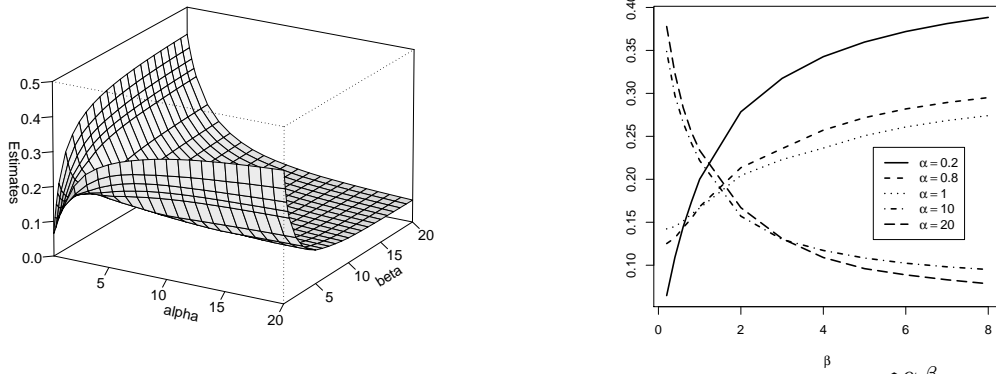


Figura 3.9: Estimativas do madograma generalizado $(\widehat{v}^{\alpha,\beta}(\mathbf{I}, \mathbf{J}))$ obtidas para $\mathbf{I} = \{\text{L.Comprida, Fajão}\}$, $\mathbf{J} = \{\text{Penamacor}\}$, com $\alpha, \beta \in \{0.2, 0.4, 0.6, 0.8, 1, 2, 3, \dots, 20\}$, à esquerda, e para $\alpha \in \{0.2, 0.8, 1, 10, 20\}$ e $\beta \in \{0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5, 6, 7, 8\}$, à direita.

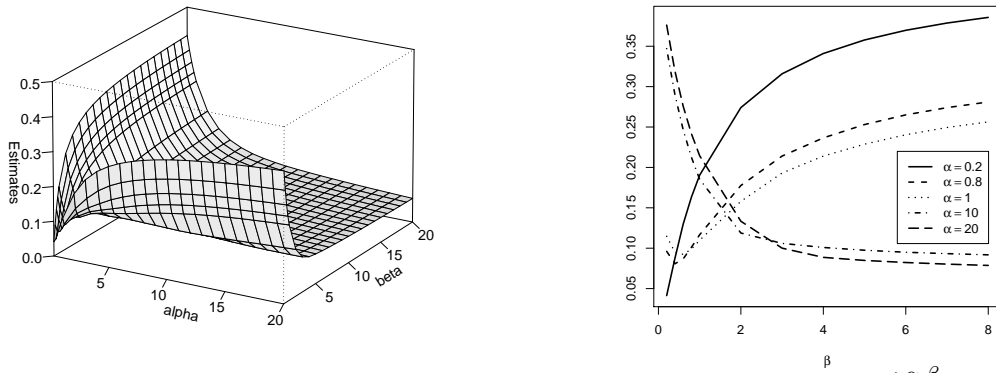


Figura 3.10: Estimativas do madograma generalizado $(\widehat{v}^{\alpha,\beta}(\mathbf{I}, \mathbf{J}))$ obtidas para $\mathbf{I} = \{\text{L.Comprida, Fajão}\}$, $\mathbf{J} = \{\text{C.Felgueira}\}$, com $\alpha, \beta \in \{0.2, 0.4, 0.6, 0.8, 1, 2, 3, \dots, 20\}$, à esquerda, e para $\alpha \in \{0.2, 0.8, 1, 10, 20\}$ e $\beta \in \{0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5, 6, 7, 8\}$, à direita.

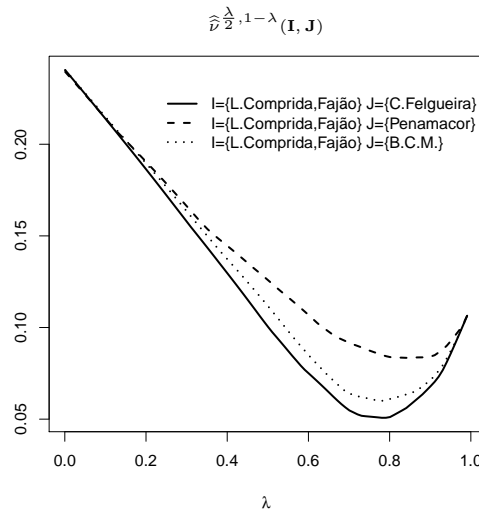


Figura 3.11: Estimativas do madograma generalizado $(\widehat{v}^{\alpha,\beta}(\mathbf{I}, \mathbf{J}))$ para $\alpha = \frac{\lambda}{2}$ e $\beta = 1 - \lambda$ com $\lambda \in (0, 1)$.

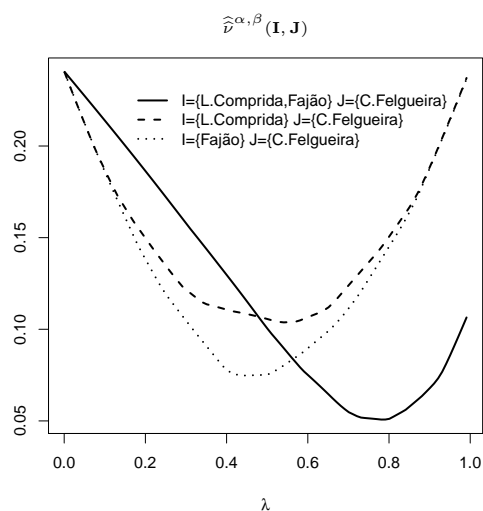


Figura 3.12: Estimativas do madograma generalizado $\widehat{v}^{\alpha, \beta}(\mathbf{I}, \mathbf{J})$ obtidas para $\mathbf{I} = \{\text{L.Comprida, Fajão}\}$, $\mathbf{J} = \{\text{C.Felgueira}\}$ com $\alpha = \frac{\lambda}{2}$ e $\beta = 1 - \lambda$, $\lambda \in (0, 1)$, e estimativas do λ -madograma para as restantes regiões.

presentados na Figura 3.11 sugerem uma dependência forte entre os níveis máximos de precipitação na região da montanha $\mathbf{I} = \{\text{Fajão, L.Comprida}\}$ e a região norte $\mathbf{J} = \{\text{C.Felgueira}\}$ sendo, também, esta a indicação dada pelos resultados apresentados nas Figuras 3.8, 3.9 e 3.10. Esta conclusão está de acordo com o que foi referido anteriormente, a Cordilheira Central é uma barreira física para a precipitação em Portugal, dificultando a deslocação da precipitação para sul.

Os valores estimados do λ -madograma para os pares de localizações (L.Comprida, C.Felgueira) e (Fajão, C.Felgueira) e os valores estimados do madograma generalizado para $\mathbf{I} = \{\text{Fajão, L.Comprida}\}$ e $\mathbf{J} = \{\text{C.Felgueira}\}$ (Figura 3.12) têm significados diferentes, já que os coeficientes correspondem a diferentes conceitos de dependência, sendo de destacar que o madograma generalizado incorpora a dependência intra-regiões e inter-regiões.

Teria sido interessante alargar este estudo a outras regiões vizinhas da Cordilheira Central, contudo os dados disponíveis não o permitiram.

Capítulo 4

Índices de contágio e estabilidade

4.1 Introdução

Na comunicação social ouve-se falar em alterações climáticas com alguma frequência, sendo este um tópico de interesse para a população em geral e, em particular, para a comunidade científica. As alterações climáticas são usualmente associadas ao aumento dos gases com efeito de estufa na atmosfera, sendo estes apontados como causadores de um aumento da temperatura global e das alterações do ciclo hidrológico (Trenberth (1998, 1999)). Trenberth também refere que é de esperar um aumento do número de ocorrências de chuvas intensas com inundações. Acontecimentos recentes, de que são exemplo as inundações no Brasil e na Austrália em 2011, demonstram que estes acontecimentos extremos têm efeitos catastróficos em termos económicos, na vida das populações (causando muitas mortes) e nos ecossistemas.

Dado o impacto destes eventos ambientais extremos, parece-nos pertinente avaliar a sua propagação a regiões vizinhas e o seu comportamento local. Por exemplo, a possibilidade de quantificar a propagação de um evento extremo de precipitação (temperatura) numa dada localização a uma região vizinha poderá auxiliar na tomada de decisão sobre a implementação, caso se justifique, de medidas preventivas nessa região de forma a minimizar os danos. Por outro lado, em termos de comportamento local é importante, por exemplo, quantificar o impacto de um valor de precipitação (temperatura) extrema na estabilidade de uma região, podendo, também neste caso, contribuir-se para desencadear uma situação de alerta na região afetada.

Neste capítulo contribuímos para a resposta a estas questões através da introdução dos coeficientes de contágio e de estabilidade (Fonseca *et al.* (2012)) em campos aleatórios max-estáveis $\mathbf{X} = \{X_i\}_{i \in \mathbb{R}^2}$ e estudo das suas propriedades, nomeadamente as suas relações com coeficientes de dependência existentes na literatura. Estas formas de caracterização dos índices motivaram os estimadores propostos na

Secção 4.5.

Com base em simulações de um campo M4, determinamos estimativas dos índices propostos e comparamo-las com os seus valores exatos. Finalizamos com uma aplicação a dados reais de precipitação anual máxima em localizações portuguesas.

4.2 Índice de contágio

O índice de contágio permitirá avaliar em que medida a ocorrência de uma excedência de um nível elevado numa localização \mathbf{i} , se poderá propagar e contagiar uma região de localizações vizinhas de \mathbf{i} . No que se segue entende-se por região de \mathbb{R}^2 qualquer subconjunto discreto $\mathbf{A} \subset \mathbb{R}^2$.

Definição 4.1 *Seja \mathbf{X} um campo aleatório max-estável com f.d. marginal univariada F e \mathbf{A} uma região de \mathbb{R}^2 . O índice de contágio da localização $\mathbf{i} \in \mathbb{R}^2$ para a região $\mathbf{A} \subset \mathbb{R}^2$ é definido por*

$$CI(\mathbf{A}, \mathbf{i}) = \lim_{u \uparrow 1} E \left(\sum_{\mathbf{j} \in \mathbf{A}} \mathbb{1}_{\{F(X_{\mathbf{j}}) > u\}} \middle| F(X_{\mathbf{i}}) > u \right), \quad (4.1)$$

desde que o limite exista.

O índice de contágio é o número esperado condicional de excedências acima de um nível elevado x em \mathbf{A} , dado que $X_{\mathbf{i}}$ excede x , ou seja, $CI(\mathbf{A}, \mathbf{i})$ mede o impacto da ocorrência do evento $\{X_{\mathbf{i}} > x\}$ na região \mathbf{A} .

Se $\mathbf{i} \notin \mathbf{A}$ tem-se $0 \leq CI(\mathbf{A}, \mathbf{i}) \leq |\mathbf{A}|$, caso contrário $1 \leq CI(\mathbf{A}, \mathbf{i}) \leq |\mathbf{A}|$. Quanto maior for o valor do índice, maior é o contágio ou influência do acontecimento $\{X_{\mathbf{i}} > x\}$ sobre a região \mathbf{A} .

No que se segue, se nada for dito em contrário, estaremos a assumir que $\mathbf{i} \notin \mathbf{A}$.

O índice de contágio da localização \mathbf{i} para a região \mathbf{A} pode ser calculado a partir dos coeficientes extremais bivariados $\varepsilon(\mathbf{i}, \mathbf{j}), \mathbf{j} \in \mathbf{A}$, definidos em (1.11), como se mostra na proposição seguinte.

Proposição 4.1 *Seja \mathbf{X} um campo aleatório max-estável com f.d. marginal univariada, F , de Fréchet unitária e $\mathbf{A} \subset \mathbb{R}^2$. Então,*

$$CI(\mathbf{A}, \mathbf{i}) = \sum_{\mathbf{j} \in \mathbf{A}} \lambda(\{\mathbf{i}\}, \{\mathbf{j}\}) = 2|\mathbf{A}| - \sum_{\mathbf{j} \in \mathbf{A}} \varepsilon(\{\mathbf{i}, \mathbf{j}\}). \quad (4.2)$$

Demonstração: O resultado é consequência imediata da definição do índice de

contágio e da relação $\lambda(\{\mathbf{i}\}, \{\mathbf{j}\}) = 2 - \varepsilon(\{\mathbf{i}, \mathbf{j}\})$. De facto,

$$\begin{aligned} CI(\mathbf{A}, \mathbf{i}) &= \sum_{\mathbf{j} \in \mathbf{A}} \lim_{u \uparrow 1} P(F(X_{\mathbf{j}}) > u | F(X_{\mathbf{i}}) > u) \\ &= \sum_{\mathbf{j} \in \mathbf{A}} \lambda(\{\mathbf{i}\}, \{\mathbf{j}\}) = 2|\mathbf{A}| - \sum_{\mathbf{j} \in \mathbf{A}} \varepsilon(\{\mathbf{i}, \mathbf{j}\}). \end{aligned}$$

□

As relações estabelecidas neste resultado evidenciam que o coeficiente de contágio pode ser usado para quantificar a dependência. Note-se que, se $X_{\mathbf{i}}$ é independente de $X_{\mathbf{j}}$, para cada $\mathbf{j} \in \mathbf{A}$, então $CI(\mathbf{A}, \mathbf{i}) = 0$. Por outro lado, se para cada $\mathbf{j} \in \mathbf{A}$, $X_{\mathbf{i}}$ e $X_{\mathbf{j}}$ são totalmente dependentes, então $CI(\mathbf{A}, \mathbf{i}) = |\mathbf{A}|$.

Coloca-se-nos agora a questão de saber como quantificar o contágio de um acontecimento extremo ocorrido numa região \mathbf{B} sobre uma região vizinha, \mathbf{A} . Para responder a esta questão, apresentamos uma extensão natural do coeficiente $CI(\mathbf{A}, \mathbf{i})$ e a sua relação com coeficientes de dependência existentes na literatura.

Definição 4.2 *Seja \mathbf{X} um campo aleatório max-estável com f.d. marginal univariada F e \mathbf{A} e \mathbf{B} regiões de \mathbb{R}^2 . O índice de contágio da região \mathbf{B} para a região \mathbf{A} define-se do seguinte modo*

$$CI(\mathbf{A}, \mathbf{B}) = \lim_{u \uparrow 1} E \left(\sum_{\mathbf{j} \in \mathbf{A}} \mathbb{1}_{\{F(X_{\mathbf{j}}) > u\}} \middle| \bigcup_{\mathbf{i} \in \mathbf{B}} \{F(X_{\mathbf{i}}) > u\} \right), \quad (4.3)$$

desde que o limite exista.

Se em (4.3) considerarmos $\mathbf{A} = \mathbf{B}$ encontramos a definição de índice de fragilidade (FI) da região \mathbf{A} . Este índice foi introduzido por Geluk *et al.* (2007) para medir a estabilidade de um sistema estocástico e estendido por Ferreira e Ferreira (2012a), onde se considera o sistema dividido em blocos. Quando $FI = 1$ o sistema diz-se estável, caso contrário diz-se frágil. Falk e Tichy (2011, 2012) estudaram a distribuição assintótica do número de excedências condicional à existência de alguma excedência.

Como consequência imediata da definição do coeficiente multivariado de dependência na cauda superior (Ferreira (2011), Li (2009), Schmidt (2002)) apresentado em (1.15), obtém-se o seguinte processo de cálculo de $CI(\mathbf{A}, \mathbf{B})$.

$$CI(\mathbf{A}, \mathbf{B}) = \lim_{u \uparrow 1} \sum_{\mathbf{j} \in \mathbf{A}} \frac{P \left(F(X_{\mathbf{j}}) > u, \bigcup_{\mathbf{i} \in \mathbf{B}} \{F(X_{\mathbf{i}}) > u\} \right)}{P \left(\bigcup_{\mathbf{i} \in \mathbf{B}} \{F(X_{\mathbf{i}}) > u\} \right)}$$

$$\begin{aligned}
 &= \lim_{u \uparrow 1} \sum_{\substack{\emptyset \neq \mathbf{J} \subseteq \mathbf{B} \\ \mathbf{j} \in \mathbf{A}}} \frac{\sum_{\mathbf{i} \in \mathbf{J}} (-1)^{|\mathbf{J}|+1} P \left(\bigcap_{\mathbf{i} \in \mathbf{J}} \{F(X_{\mathbf{i}}) > u, F(X_{\mathbf{j}}) > u\} \right)}{1 - u^{\varepsilon(\mathbf{B})}} \\
 &= \sum_{\substack{\emptyset \neq \mathbf{J} \subseteq \mathbf{B} \\ \mathbf{j} \in \mathbf{A}}} \frac{\sum_{\mathbf{i} \in \mathbf{J}} (-1)^{|\mathbf{J}|+1} \lambda(\mathbf{J}, \{\mathbf{j}\})}{\varepsilon(\mathbf{B})}.
 \end{aligned}$$

Observação 4.1 *A ocorrência de alguma excedência de um nível elevado pelas variáveis do campo indexadas numa região $\mathbf{B} \subseteq \mathbb{R}^2$ pode ser considerada um acontecimento extremo. Todavia, poderão existir outros de interesse, como por exemplo $\bigcap_{\mathbf{i} \in \mathbf{B}} \{X_{\mathbf{i}} > x\}$, com x elevado. O contágio deste acontecimento extremo sobre uma região \mathbf{A} pode ser quantificado do seguinte modo,*

$$CI^*(\mathbf{A}, \mathbf{B}) = \lim_{u \uparrow 1} E \left(\sum_{\mathbf{j} \in \mathbf{A}} \mathbb{1}_{\{F(X_{\mathbf{j}}) > u\}} \left| \bigcap_{\mathbf{i} \in \mathbf{B}} \{F(X_{\mathbf{i}}) > u\} \right. \right),$$

e relaciona-se com o coeficiente médio de dependência na cauda apresentado em (2.11) da seguinte forma

$$CI^*(\mathbf{A}, \mathbf{B}) = |\mathbf{A}| \bar{\lambda}(\mathbf{A}, \mathbf{B}).$$

4.3 Índice de estabilidade

Nesta secção introduz-se um novo coeficiente de dependência espacial, o índice de estabilidade, para analisar a oscilação de \mathbf{X} numa região \mathbf{A} relativamente a uma localização específica \mathbf{i} .

Definição 4.3 *Seja \mathbf{X} um campo aleatório max-estável com f.d. marginal univariada F e \mathbf{A} uma região de \mathbb{R}^2 . O índice de estabilidade da região \mathbf{A} associado a uma localização específica $\mathbf{i} \in \mathbb{R}^2$, $SI(\mathbf{A}, \mathbf{i})$, é definido por*

$$SI(\mathbf{A}, \mathbf{i}) = \lim_{u \uparrow 1} E \left(\sum_{\mathbf{j} \in \mathbf{A}} \mathbb{1}_{\{F(X_{\mathbf{i}}) \leq u < F(X_{\mathbf{j}})\}} \left| \sum_{\mathbf{j} \in \mathbf{A}} \mathbb{1}_{\{F(X_{\mathbf{i}}) \leq u < F(X_{\mathbf{j}})\}} > 0 \right. \right),$$

desde que o limite exista.

Designando por cruzamento do nível x a partir da localização \mathbf{i} qualquer acontecimento do tipo $\{X_{\mathbf{i}} \leq x \leq X_{\mathbf{j}}\}$, para $\mathbf{j} \in \mathbb{R}^2 \setminus \{\mathbf{i}\}$, o índice de estabilidade é dado pelo número esperado condicional de cruzamentos de um nível elevado x em \mathbf{A} a

partir da localização \mathbf{i} , dado que existe pelo menos um cruzamento de x em \mathbf{A} a partir de \mathbf{i} .

Se a variação de \mathbf{X} numa região de localizações \mathbf{A} não é suave, relativamente a uma localização \mathbf{i} , poderemos esperar um elevado número de cruzamentos de um nível elevado x nessa região a partir da localização \mathbf{i} . Um número elevado de cruzamentos é indicador de instabilidade na região em causa.

O próximo resultado estabelece relações entre o índice de estabilidade e os coeficientes extremais bivariado (1.11) e multivariado (1.12) e por conseguinte com o índice de contágio.

Proposição 4.2 *Seja \mathbf{X} um campo aleatório max-estável com f.d. marginal univariada, F , de Fréchet unitária, $\mathbf{i} \in \mathbb{R}^2$ e $\mathbf{A} \subset \mathbb{R}^2$. Então,*

$$SI(\mathbf{A}, \mathbf{i}) = \frac{|\mathbf{A}| - CI(\mathbf{A}, \mathbf{i})}{\varepsilon(\{\mathbf{i}\} \cup \mathbf{A}) - 1}. \quad (4.4)$$

Demonstração: Como consequência imediata da definição do coeficiente extremal, obtemos

$$\begin{aligned} SI(\mathbf{A}, \mathbf{i}) &= \lim_{u \uparrow 1} \frac{\sum_{\mathbf{j} \in \mathbf{A}} P(F(X_{\mathbf{i}}) \leq u < F(X_{\mathbf{j}}))}{P\left(F(X_{\mathbf{i}}) \leq u, \bigcup_{\mathbf{j} \in \mathbf{A}} \{F(X_{\mathbf{j}}) > u\}\right)} \\ &= \lim_{u \uparrow 1} \frac{\sum_{\mathbf{j} \in \mathbf{A}} P(F(X_{\mathbf{i}}) \leq u) - \sum_{\mathbf{j} \in \mathbf{A}} P(F(X_{\mathbf{i}}) \leq u, F(X_{\mathbf{j}}) \leq u)}{P(F(X_{\mathbf{i}}) \leq u) - P\left(F(X_{\mathbf{i}}) \leq u, \bigcap_{\mathbf{j} \in \mathbf{A}} \{F(X_{\mathbf{j}}) \leq u\}\right)} \\ &= \lim_{u \uparrow 1} \frac{|\mathbf{A}| u - \sum_{\mathbf{j} \in \mathbf{A}} u^{\varepsilon(\{\mathbf{i}, \mathbf{j}\})}}{u - u^{\varepsilon(\{\mathbf{i}\} \cup \mathbf{A})}} \\ &= \frac{\sum_{\mathbf{j} \in \mathbf{A}} \varepsilon(\{\mathbf{i}, \mathbf{j}\}) - |\mathbf{A}|}{\varepsilon(\{\mathbf{i}\} \cup \mathbf{A}) - 1} \\ &= \frac{|\mathbf{A}| - CI(\mathbf{A}, \mathbf{i})}{\varepsilon(\{\mathbf{i}\} \cup \mathbf{A}) - 1}. \end{aligned} \quad (4.5)$$

□

Observação 4.2 *Se as variáveis $X_{\mathbf{i}}, X_{\mathbf{i}_1}, \dots, X_{\mathbf{i}_k}$, forem totalmente dependentes o índice de estabilidade não está definido. Tal é natural uma vez que nestas condições se tem, para todo $\mathbf{j} \in \{\mathbf{i}_1, \dots, \mathbf{i}_k\}$,*

$$P(F(X_{\mathbf{i}}) \leq u < F(X_{\mathbf{j}})) = P(F(X_{\mathbf{i}}) \leq u) - P^{\varepsilon(\{\mathbf{i}, \mathbf{j}\})}(F(X_{\mathbf{i}}) \leq u) = 0.$$

No caso das variáveis serem independentes obtemos $SI(\mathbf{A}, \mathbf{i}) = 1$.

Tendo em conta que

$$\bigvee_{\mathbf{j} \in \mathbf{A}} \varepsilon(\{\mathbf{i}, \mathbf{j}\}) \leq \varepsilon(\{\mathbf{i}\} \cup \mathbf{A}) \leq |\mathbf{A}| + 1,$$

obtém-se o seguinte enquadramento para o índice de estabilidade.

Proposição 4.3 *Nas condições da Proposição 4.2, tem-se*

$$\frac{\sum_{\mathbf{j} \in \mathbf{A}} \varepsilon(\{\mathbf{i}, \mathbf{j}\}) - |\mathbf{A}|}{|\mathbf{A}|} \leq SI(\mathbf{A}, \mathbf{i}) \leq \frac{\sum_{\mathbf{j} \in \mathbf{A}} \varepsilon(\{\mathbf{i}, \mathbf{j}\}) - |\mathbf{A}|}{\bigvee_{\mathbf{j} \in \mathbf{A}} \varepsilon(\{\mathbf{i}, \mathbf{j}\}) - 1}.$$

O intervalo de variação do índice de estabilidade apresentado na proposição anterior é o de menor amplitude possível, como poderemos constatar nos Exemplos 4.1 e 4.2 da secção seguinte.

4.4 Exemplos

Nesta secção aplicamos os índices introduzidos ao campo aleatório M4 definido em (2.20). Como consequência imediata do coeficiente extremal do campo aleatório M4, dado por (2.23), e das Proposições 4.1 e 4.2 obtêm-se as seguintes expressões dos índices de contágio e de estabilidade

$$CI(\mathbf{A}, \mathbf{i}) = 2|\mathbf{A}| - \sum_{\mathbf{j} \in \mathbf{A}} \sum_{l=1}^{+\infty} \sum_{m=-\infty}^{+\infty} (a_{lmi} \vee a_{lmj}) \quad (4.6)$$

e

$$SI(\mathbf{A}, \mathbf{i}) = \frac{\sum_{\mathbf{j} \in \mathbf{A}} \sum_{l=1}^{+\infty} \sum_{m=-\infty}^{+\infty} (a_{lmi} \vee a_{lmj}) - |\mathbf{A}|}{\sum_{l=1}^{+\infty} \sum_{m=-\infty}^{+\infty} \left(\bigvee_{\mathbf{j} \in \mathbf{A}} a_{lmj} \vee a_{lmi} \right) - 1}. \quad (4.7)$$

Os exemplos que se seguem ilustram as situações extremas de independência e total dependência e mostram que o intervalo de variação do índice de estabilidade obtido na Proposição 4.3 é o de menor amplitude possível.

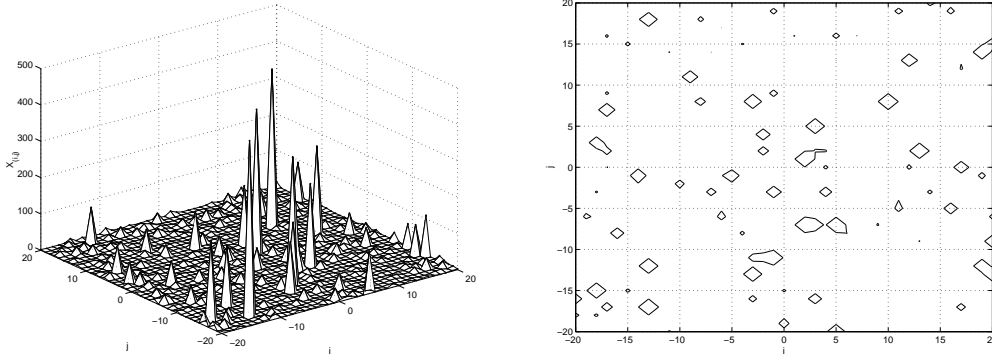


Figura 4.1: Simulação do campo aleatório M4 definido no Exemplo 4.1 (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 16.4669$ (direita).

Exemplo 4.1 Considere-se o campo aleatório M4 (Figura 4.1) com $L = 1$, $1 \leq m \leq 4$ e para cada localização $\mathbf{i} = (i_1, i_2) \in \mathbb{Z}^2$ as seguintes constantes

$$\left\{ \begin{array}{l} a_{11\mathbf{i}} = 1 \wedge a_{12\mathbf{i}} = a_{13\mathbf{i}} = a_{14\mathbf{i}} = 0 \quad \text{se } i_1 \text{ e } i_2 \text{ são pares} \\ a_{12\mathbf{i}} = 1 \wedge a_{11\mathbf{i}} = a_{13\mathbf{i}} = a_{14\mathbf{i}} = 0 \quad \text{se } i_1 \text{ é par e } i_2 \text{ ímpar} \\ a_{13\mathbf{i}} = 1 \wedge a_{11\mathbf{i}} = a_{12\mathbf{i}} = a_{14\mathbf{i}} = 0 \quad \text{se } i_1 \text{ é ímpar e } i_2 \text{ par} \\ a_{14\mathbf{i}} = 1 \wedge a_{11\mathbf{i}} = a_{12\mathbf{i}} = a_{13\mathbf{i}} = 0 \quad \text{se } i_1 \text{ e } i_2 \text{ são ímpares.} \end{array} \right.$$

Seja $\mathbf{A} = \{(1, 1), (1, 2), (2, 2)\}$ e $\mathbf{i} = (4, 1)$. Atendendo a (2.23) obtém-se, para todo $\mathbf{j} \in \mathbf{A}$, $\varepsilon(\mathbf{i}, \mathbf{j}) = 2$ e conseqüentemente, por (4.6), $CI(\mathbf{A}, \mathbf{i}) = 0$, mostrando que a ocorrência de uma excedência de um nível elevado na localização \mathbf{i} não se propagará à região \mathbf{A} .

No que diz respeito ao índice de estabilidade, aplicando (4.7) obtemos $SI(\mathbf{A}, \mathbf{i}) = 1$, valor coincidente com o limite inferior do intervalo de variação apresentado na Proposição 4.3.

Exemplo 4.2 Consideremos, neste caso, o campo M4 com $L = 1$, $1 \leq m \leq 2$ e para cada localização $\mathbf{i} = (i_1, i_2) \in \mathbb{Z}^2$ o seguinte padrão de assinaturas

$$\left\{ \begin{array}{l} a_{11\mathbf{i}} = \frac{4}{5} \wedge a_{12\mathbf{i}} = \frac{1}{5} \quad \text{se } i_1 \text{ é par} \\ a_{11\mathbf{i}} = \frac{1}{4} = 1 - a_{12\mathbf{i}} \quad \text{se } i_1 \text{ é ímpar.} \end{array} \right.$$

Seja $\mathbf{A}^{(k)} = \{s_j^k(\mathbf{i}) : \mathbf{i} = (3, 3) \wedge j \in \{1, 2, \dots, 8\}\}$, $k \in \mathbb{N}$, onde para cada $j \in \{1, \dots, 8\}$, $s_j^k(\mathbf{i})$ denota o vizinho de \mathbf{i} segundo a direção j , a passo k , definido na Secção 2.2.

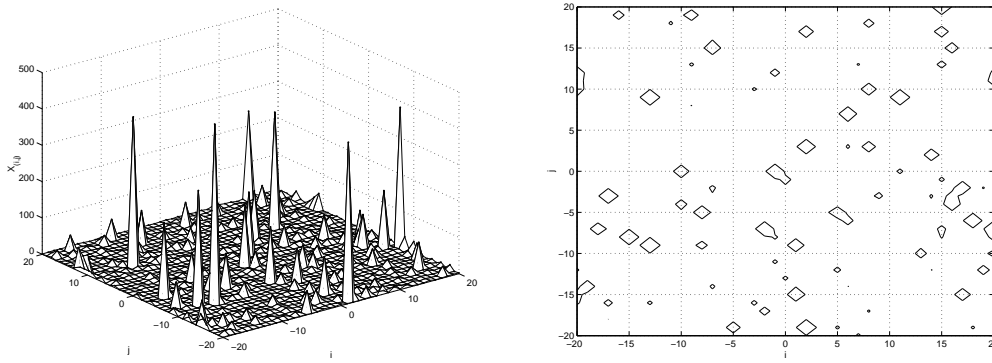


Figura 4.2: Simulação do campo aleatório M4 definido no Exemplo 4.2 (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 19.2996$ (direita).

Embora a matriz dos coeficientes extremais bivariados, $\varepsilon(s_j(\mathbf{i}), \mathbf{i})$, $j \in \{1, \dots, 8\}$, dada por

$$\begin{bmatrix} \varepsilon(s_4^k(\mathbf{i}), \mathbf{i}) & \varepsilon(s_3^k(\mathbf{i}), \mathbf{i}) & \varepsilon(s_2^k(\mathbf{i}), \mathbf{i}) \\ \varepsilon(s_5^k(\mathbf{i}), \mathbf{i}) & \varepsilon(\mathbf{i}, \mathbf{i}) & \varepsilon(s_1^k(\mathbf{i}), \mathbf{i}) \\ \varepsilon(s_6^k(\mathbf{i}), \mathbf{i}) & \varepsilon(s_7^k(\mathbf{i}), \mathbf{i}) & \varepsilon(s_8^k(\mathbf{i}), \mathbf{i}) \end{bmatrix} = \begin{cases} \begin{bmatrix} 31/20 & 1 & 31/20 \\ 31/20 & 1 & 31/20 \\ 31/20 & 1 & 31/20 \end{bmatrix} & \text{se } k \text{ é ímpar} \\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} & \text{se } k \text{ é par} \end{cases}$$

permita conhecer a probabilidade de contágio de $\mathbf{i} = (3, 3)$ aos seus vizinhos $s_j(\mathbf{i})$, $j \in \{1, \dots, 8\}$, não permite avaliar a dimensão do impacto. Tal é conseguido com o coeficiente de contágio, que é dado por

$$CI(\mathbf{A}^{(k)}, \mathbf{i}) = \begin{cases} 4.7 & \text{se } k \text{ é ímpar} \\ 8 & \text{se } k \text{ é par.} \end{cases}$$

Quando k é ímpar o índice de estabilidade da região $\mathbf{A}^{(k)}$ associado à localização \mathbf{i} coincide com o limite superior do intervalo de variação apresentado na Proposição 4.3 e é igual a 6, sendo indicador de forte instabilidade do campo na região $\mathbf{A}^{(k)}$, relativamente à localização \mathbf{i} . No caso em que k é par tem-se a total dependência, por conseguinte $SI(\mathbf{A}^{(k)}, \mathbf{i})$ não está definido.

Estes resultados são um reflexo da definição do padrão de assinaturas, que depende da paridade da abscissa das localizações, por este determinar a estrutura de dependência do campo M4.

Nos exemplos que apresentámos até ao momento, ilustrámos as situações extremas dos dois índices em campos aleatórios M4 com um padrão de assinaturas e, por isso, com uma estrutura de dependência simples. No exemplo que se segue va-

mos considerar uma estrutura de dependência um pouco mais complexa, um campo aleatório M4 com quatro padrões de assinaturas.

Exemplo 4.3 Considere-se um campo aleatório M4, $\mathbf{X} = \{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathbb{Z}^2}$, com $L = 4$ e $1 \leq m \leq 2$ (Figura 4.3) definido por quatro padrões de assinaturas da forma $(a_{l1\mathbf{i}}, a_{l2\mathbf{i}})$, $l \in \{1, 2, 3, 4\}$, sendo as constantes associadas a cada localização $\mathbf{i} = (i_1, j_1) \in \mathbb{Z}^2$ as seguintes

- Se $i_1 > j_1 \wedge i_1$ par $\wedge j_1$ ímpar, então

$$a_{11\mathbf{i}} = a_{12\mathbf{i}} = a_{21\mathbf{i}} = a_{22\mathbf{i}} = \frac{1}{8}$$

$$a_{31\mathbf{i}} = a_{32\mathbf{i}} = a_{41\mathbf{i}} = a_{42\mathbf{i}} = \frac{1}{8}$$
- Se $i_1 \leq j_1 \wedge i_1$ par $\wedge j_1$ ímpar, então

$$a_{11\mathbf{i}} = a_{12\mathbf{i}} = \frac{2}{17}, a_{21\mathbf{i}} = \frac{5}{17}, a_{22\mathbf{i}} = \frac{4}{17}$$

$$a_{31\mathbf{i}} = a_{32\mathbf{i}} = a_{41\mathbf{i}} = a_{42\mathbf{i}} = \frac{1}{17}$$
- Se $i_1 > j_1 \wedge i_1$ ímpar $\wedge j_1$ par, então

$$a_{11\mathbf{i}} = \frac{1}{20}, a_{12\mathbf{i}} = \frac{2}{20}, a_{21\mathbf{i}} = \frac{3}{20}, a_{22\mathbf{i}} = \frac{4}{20}$$

$$a_{31\mathbf{i}} = \frac{5}{20}, a_{32\mathbf{i}} = \frac{3}{20}, a_{41\mathbf{i}} = a_{42\mathbf{i}} = \frac{1}{20}$$
- Se $i_1 \leq j_1 \wedge i_1$ ímpar $\wedge j_1$ par, então

$$a_{11\mathbf{i}} = \frac{1}{36}, a_{12\mathbf{i}} = \frac{2}{36}, a_{21\mathbf{i}} = \frac{3}{36}, a_{22\mathbf{i}} = \frac{4}{36}$$

$$a_{31\mathbf{i}} = \frac{5}{36}, a_{32\mathbf{i}} = \frac{6}{36}, a_{41\mathbf{i}} = \frac{7}{36}, a_{42\mathbf{i}} = \frac{8}{36}$$
- Se $i_1 > j_1 \wedge i_1$ par $\wedge j_1$ par, então

$$a_{11\mathbf{i}} = \frac{1}{40}, a_{12\mathbf{i}} = \frac{2}{40}, a_{21\mathbf{i}} = \frac{3}{40}, a_{22\mathbf{i}} = \frac{4}{40}$$

$$a_{31\mathbf{i}} = \frac{5}{40}, a_{32\mathbf{i}} = \frac{6}{40}, a_{41\mathbf{i}} = \frac{7}{40}, a_{42\mathbf{i}} = \frac{12}{40}$$
- Se $i_1 \leq j_1 \wedge i_1$ par $\wedge j_1$ par, então

$$a_{11\mathbf{i}} = \frac{1}{45}, a_{12\mathbf{i}} = \frac{2}{45}, a_{21\mathbf{i}} = \frac{3}{45}, a_{22\mathbf{i}} = \frac{4}{45}$$

$$a_{31\mathbf{i}} = \frac{6}{45}, a_{32\mathbf{i}} = \frac{8}{45}, a_{41\mathbf{i}} = \frac{9}{45}, a_{42\mathbf{i}} = \frac{12}{45}$$
- Se $i_1 > j_1 \wedge i_1$ ímpar $\wedge j_1$ ímpar, então

$$a_{11\mathbf{i}} = \frac{1}{50}, a_{12\mathbf{i}} = \frac{7}{50}, a_{21\mathbf{i}} = \frac{3}{50}, a_{22\mathbf{i}} = \frac{4}{50}$$

$$a_{31\mathbf{i}} = \frac{6}{50}, a_{32\mathbf{i}} = \frac{8}{50}, a_{41\mathbf{i}} = \frac{9}{50}, a_{42\mathbf{i}} = \frac{12}{50}$$
- Se $i_1 \leq j_1 \wedge i_1$ ímpar $\wedge j_1$ ímpar, então

$$a_{11\mathbf{i}} = \frac{1}{60}, a_{12\mathbf{i}} = \frac{7}{60}, a_{21\mathbf{i}} = \frac{3}{60}, a_{22\mathbf{i}} = \frac{14}{60}$$

$$a_{31\mathbf{i}} = \frac{6}{60}, a_{32\mathbf{i}} = \frac{8}{60}, a_{41\mathbf{i}} = \frac{9}{60}, a_{42\mathbf{i}} = \frac{12}{60}.$$

Considerando $\mathbf{i} = (3, 3)$, $\mathbf{A} = \{(4, 3), (3, 4), (2, 3), (2, 2), (3, 2), (4, 2)\}$ e aplicando (2.23) obtém-se

$$\varepsilon(\{(4, 3), \mathbf{i}\}) = \frac{73}{60}, \quad \varepsilon(\{(3, 4), \mathbf{i}\}) = \frac{71}{60}, \quad \varepsilon(\{(2, 3), \mathbf{i}\}) = \frac{275}{204}$$

$$\varepsilon(\{(2, 2), \mathbf{i}\}) = \frac{73}{60}, \quad \varepsilon(\{(3, 2), \mathbf{i}\}) = \frac{13}{10}, \quad \varepsilon(\{(4, 2), \mathbf{i}\}) = \frac{6}{5}$$

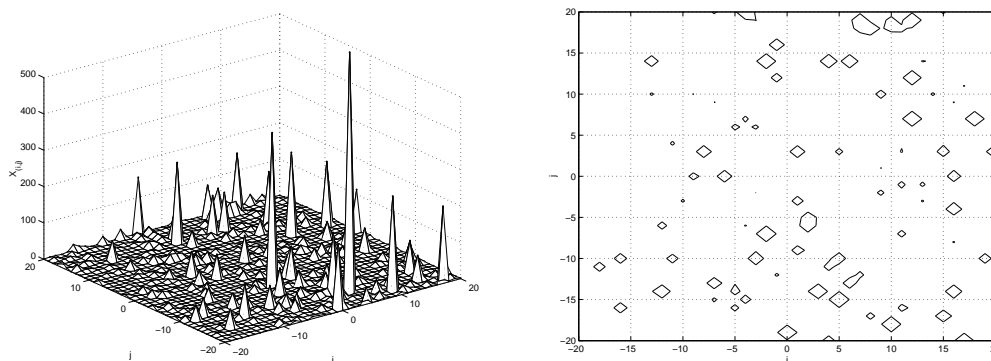


Figura 4.3: Simulação do campo aleatório M4 definido no Exemplo 4.3 (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 19.5652$ (direita).

e por conseguinte, $CI(\mathbf{A}, \mathbf{i}) \approx 4.5353$ e $SI(\mathbf{A}, \mathbf{i}) \approx 2.0712$. O valor do índice de contágio indica que a propagação de uma excedência de um nível elevado em \mathbf{i} à região \mathbf{A} é significativa. O valor do índice de estabilidade indica que o número esperado de cruzamentos em \mathbf{A} , a partir de \mathbf{i} , será aproximadamente igual a 2.

4.5 Estimação

As relações estabelecidas nas Proposições 4.1 e 4.2, permitem estimar de imediato os índices de contágio e estabilidade, através da estimação do coeficiente extremal $\varepsilon(\{\mathbf{i}\} \cup \mathbf{A})$, $\mathbf{A} \subset \mathbb{R}^2$. Mais precisamente, os estimadores de $CI(\mathbf{A}, \mathbf{i})$ e $SI(\mathbf{A}, \mathbf{i})$ definem-se por

$$\widehat{CI}(\mathbf{A}, \mathbf{i}) = 2|\mathbf{A}| - \sum_{\mathbf{j} \in \mathbf{A}} \widehat{\varepsilon}(\{\mathbf{i}, \mathbf{j}\}) \quad (4.8)$$

e

$$\widehat{SI}(\mathbf{A}, \mathbf{i}) = \frac{\sum_{\mathbf{j} \in \mathbf{A}} \widehat{\varepsilon}(\{\mathbf{i}, \mathbf{j}\}) - |\mathbf{A}|}{\widehat{\varepsilon}(\{\mathbf{i}\} \cup \mathbf{A}) - 1}. \quad (4.9)$$

Autores como Einmahl (2012), Krajina (2010) e Beirlant (2004) apresentam propostas de estimadores para a função de dependência na cauda dada em (3.4), que podem ser utilizados na estimação de $\varepsilon(\{\mathbf{i}\} \cup \mathbf{A})$, uma vez que

$$\varepsilon(\{\mathbf{i}\} \cup \mathbf{A}) = l_{\mathbf{A} \cup \{\mathbf{i}\}}(\mathbf{1}),$$

onde $\mathbf{1}$ denota o vetor unitário $(|\mathbf{A}| + 1)$ -dimensional.

A abordagem utilizada por qualquer um destes autores envolve a escolha de um valor k que corresponde ao número de observações de topo que deverá ser utilizado na estimação de um quantil elevado, conduzindo ao problema usual da variância e do viés, tal como tivemos oportunidade de referir na Secção 3.3. Para ultrapassar esta limitação, em Ferreira e Ferreira (2012b) propõe-se uma outra abordagem para

n	\widehat{CI}	MSE
100	8	1.64e-29
500	8	6.04e-29
1000	8	1.18e-28

Tabela 4.1: Média dos valores estimados do índice de contágio, \widehat{CI} , e erro quadrático médio, MSE , para o Exemplo 4.2 onde $CI(\mathbf{A}^{(k)}, \mathbf{i}) = 8$, k par.

estimar o coeficiente extremal e que não depende da escolha de k .

O referido estimador baseia-se na relação

$$\varepsilon(\mathbf{A}) = V_{\mathbf{A}}(1, 1, \dots, 1) = \frac{E(M(\mathbf{A}))}{1 - E(M(\mathbf{A}))}, \quad \text{onde } M(\mathbf{A}) = \bigvee_{\mathbf{i} \in \mathbf{A}} F_{\mathbf{i}}(X_{\mathbf{i}}).$$

Considerando o correspondente empírico do valor esperado, o estimador é definido por

$$\widehat{\varepsilon}(\mathbf{A}) = \frac{\overline{M(\mathbf{A})}}{1 - \overline{M(\mathbf{A})}}, \quad (4.10)$$

onde

$$\overline{M(\mathbf{A})} = \frac{1}{n} \sum_{j=1}^n \bigvee_{\mathbf{i} \in \mathbf{A}} F_{\mathbf{i}}(X_{\mathbf{i}}^{(j)}) \quad (4.11)$$

sendo $F_{\mathbf{i}}$, $\mathbf{i} \in \mathbf{A}$, a f.d. de $X_{\mathbf{i}}$. No caso em que $F_{\mathbf{i}}$ é desconhecida as autoras utilizam a correspondente função de distribuição empírica modificada, apresentada em (3.9), e provam a normalidade assintótica e a consistência do estimador, provando também estes resultados quando a f.d. marginal univariada é conhecida.

No que se segue consideramos a abordagem de Ferreira e Ferreira (2012b) na definição dos estimadores $\widehat{CI}(\mathbf{A}, \mathbf{i})$ e $\widehat{SI}(\mathbf{A}, \mathbf{i})$. A normalidade assintótica e a consistência destes estimadores é consequência imediata da normalidade assintótica e da consistência do estimador do coeficiente extremal proposto por Ferreira e Ferreira (2012b).

Com o objetivo de avaliar o desempenho dos estimadores efetuaram-se 100 réplicas, para amostras de dimensão $n \in \{100, 500, 1000\}$ em cada localização, de cada um dos modelos considerados nos Exemplos 4.1, 4.2 e 4.3.

Observando as Tabelas 4.1, 4.2, 4.3 e 4.4, as quais resumem os resultados obtidos, podemos constatar que os valores estimados dos coeficientes de contágio e de estabilidade se encontram, de um modo geral, bastante próximos dos respectivos verdadeiros valores.

O bom desempenho dos estimadores não paramétricos \widehat{CI} e \widehat{SI} tornam-nos ferramentas úteis na avaliação, em campos aleatórios max-estáveis, do efeito de contágio e da regularidade em regiões de localizações.

n	I. de contágio		I. de estabilidade	
	\widehat{CI}	MSE	\widehat{SI}	MSE
100	0.1282	0.0250	0.9985	0.0022
500	0.0529	0.0049	0.9959	0.0003
1000	0.0429	0.0033	0.9997	0.0002

Tabela 4.2: Média dos valores estimados dos índices de contágio, \widehat{CI} , e de estabilidade, \widehat{SI} , e do erro quadrático médio, MSE , para o Exemplo 4.1 onde $CI(\mathbf{A}, \mathbf{i}) = 0$ e $SI(\mathbf{A}, \mathbf{i}) = 1$.

n	I. de contágio		I. de estabilidade	
	\widehat{CI}	MSE	\widehat{SI}	MSE
100	4.7848	0.1232	6	3.60e-29
500	4.7449	0.0270	6	1.97e-29
1000	4.7081	0.0126	6	1.97e-29

Tabela 4.3: Média dos valores estimados dos índices de contágio, \widehat{CI} , e de estabilidade, \widehat{SI} , e do erro quadrático médio, MSE , para o Exemplo 4.2 onde $CI(\mathbf{A}^{(k)}, \mathbf{i}) = 4.7$ e $SI(\mathbf{A}^{(k)}, \mathbf{i}) = 6$, k ímpar.

n	I. de contágio		I. de estabilidade	
	\widehat{CI}	MSE	\widehat{SI}	MSE
100	4.5463	0.0202	2.0451	0.0178
500	4.5309	0.0047	2.0606	0.0033
1000	4.5339	0.0021	2.0721	0.0015

Tabela 4.4: Média dos valores estimados dos índices de contágio, \widehat{CI} , e de estabilidade, \widehat{SI} , e do erro quadrático médio, MSE , para o Exemplo 4.3 onde $CI(\mathbf{A}, \mathbf{i}) \approx 4.5353$ e $SI(\mathbf{A}, \mathbf{i}) \approx 2.0712$.

4.6. APLICAÇÃO A DADOS DE PRECIPITAÇÃO

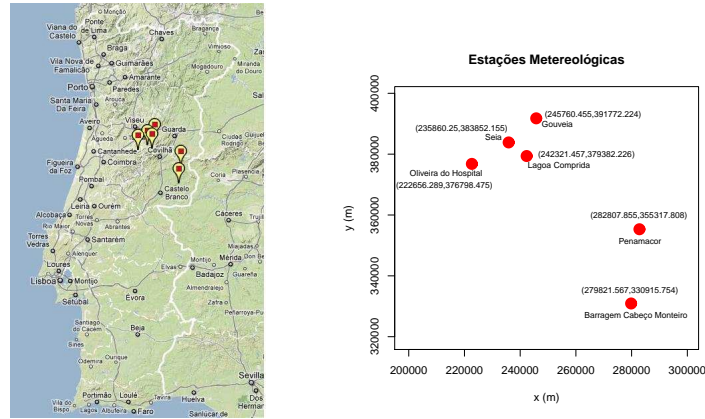


Figura 4.4: Localização das estações, obtida no SNIRH (esquerda), e a sua representação em coordenadas Lambert (direita).

A	$\widehat{CI}(A, \text{Lagoa Comprida})$	$\widehat{SI}(A, \text{Lagoa Comprida})$
{Gouveia, Oliveira do Hospital, Seia}	0.96688	1.601702
{Penamacor, Barragem Cabeço Monteiro}	0.008872	1.222402

Tabela 4.5: Estimativas dos índices de contágio e de estabilidade nas regiões {Gouveia, Oliveira do Hospital, Seia} e {Penamacor, Barragem Cabeço Monteiro} relativamente a “Lagoa Comprida”.

4.6 Aplicação a dados de precipitação

Nesta secção apresentam-se os resultados da aplicação dos estimadores dos índices de contágio e estabilidade a dados de precipitação máxima anual obtidos de valores máximos diários registados num período de 32 anos (1944-1949, 1951-1952, 1954-1971, 1974-1975, 1977, 1979-1981), em seis estações portuguesas Figura 4.4, que estão disponíveis no SNIRH e algumas das quais já consideradas na Secção 3.5.

Sendo os dados máximos relativos a um longo período temporal assume-se que, para cada localização, são independentes ao longo dos anos. Considera-se, ainda, que o campo aleatório é max-estável com distribuição marginal univariada desconhecida.

Na Tabela 4.5 apresentam-se as estimativas dos índices de contágio e estabilidade sobre as regiões {Gouveia, Oliveira do Hospital, Seia} e {Penamacor, Barragem Cabeço Monteiro} relativamente a “Lagoa Comprida” (ponto com maior altitude).

Os resultados sugerem que em termos de volume de precipitação, “Lagoa Comprida” tem uma maior influência na região {Gouveia, Oliveira do Hospital, Seia} sendo esta mais instável do que a região {Penamacor, Barragem Cabeço Monteiro}.

A região {Gouveia, Oliveira do Hospital, Seia}, está situada a noroeste da Cordilheira Central (SO-NE) e pouco distante desta, o que poderá justificar o maior volume de precipitação. Por outro lado a região {Penamacor, Barragem Cabeço Monteiro} está afastada e a sul da cordilheira, que exerce o efeito de barreira, pelo

CAPÍTULO 4. ÍNDICES DE CONTÁGIO E ESTABILIDADE

que o contágio desta região pela "Lagoa Comprida" é pouco significativo.

Capítulo 5

Um modelo autoregressivo de máximos

5.1 Introdução

É frequente ajustarem-se sucessões Markovianas com margens não simétricas e não estáveis ao estudo de máximos e excedências em hidrologia e noutros fenómenos geofísicos. Assim, Alpuim (1989) propôs a sucessão Markoviana definida através da relação

$$X_i = k(X_{i-1} \vee Y_i), \quad i \geq 1, \quad 0 \leq k \leq 1, \quad (5.1)$$

onde $\{Y_n\}_{n \geq 1}$ é uma sucessão de v.a.'s i.i.d. e também independentes da v.a. X_0 .

Esta sucessão apresenta propriedades particularmente interessantes no que respeita ao comportamento extremal (Alpuim (1989)) e corresponde ao caso particular $\alpha = 0$ do modelo

$$X_i = \beta X_{i-1} \vee (\alpha \beta X_{i-1} + Y_i), \quad i \geq 1, \quad 0 \leq \alpha \leq 1, \quad 0 < \beta < 1,$$

introduzido por Haslett (1979) para descrever o nível de temperatura num tanque utilizado para armazenar energia solar e, posteriormente, estudado por Daley e Haslett (1982), Greenwood e Hooghiemstra (1988), entre outros.

Inspirados no modelo de Haslett (1979) e com o objetivo de incorporar a componente espacial no fenómeno em estudo, propomos (Fonseca *et al.* (2010)) um campo autoregressivo de máximos (ARMAX) construído a partir de uma sucessão estacionária $\mathbf{X}^{(0)} = \{X_{(i,0)}\}_{i \geq 1}$ e de sucessões $\mathbf{Y}^{(j)} = \{Y_{(i,j)}\}_{i \geq 1}$, $j \in \mathbb{N}$, estacionárias, independentes entre si e de $\mathbf{X}^{(0)}$, e identicamente distribuídas.

Definimos o campo ARMAX $\mathbf{X} = \{X_{(i,j)}\}_{(i,j) \in \mathbb{N}^2}$, através da relação

$$X_{(i,j)} = k(X_{(i,j-1)} \vee Y_{(i,j)}), \quad (i,j) \in \mathbb{N}^2, \quad 0 < k < 1, \quad (5.2)$$

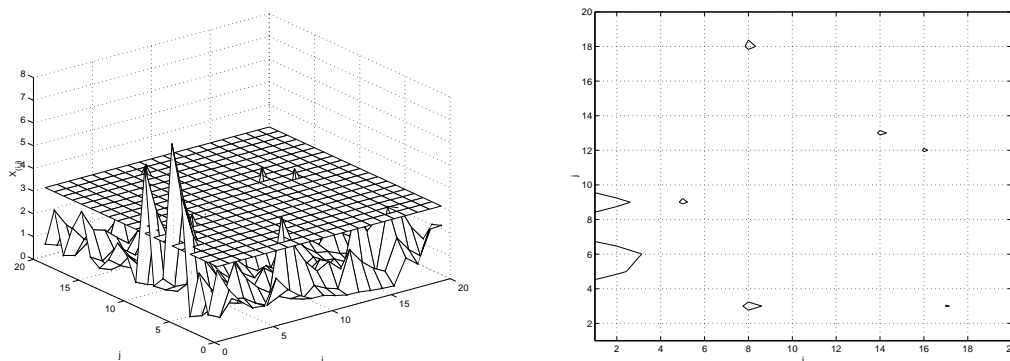


Figura 5.1: Simulação do campo aleatório ARMAX com $k = 0.65$ (esquerda) e representação das curvas de nível para o percentil 95, $x_{(i,j)} = 3.0174$ (direita).

onde, para cada valor fixo $i \geq 1$, obtém-se a sucessão Markoviana (5.1) estudada por Alpuim.

Na Secção 5.2 apresentamos algumas fórmulas e propriedades básicas do campo ARMAX. O estudo da estacionariedade e da max-estabilidade desse campo será abordado nas Secções 5.3 e 5.4, respetivamente. Finalmente, na Secção 5.5, faremos uma análise da dependência no campo ARMAX, com base nos coeficientes propostos nos capítulos anteriores.

5.2 Algumas fórmulas básicas

Seja $\mathbf{X} = \{X_{(i,j)}\}_{(i,j) \in \mathbb{N}^2}$, o campo ARMAX definido em (5.2). Na Figura 5.1 apresenta-se uma simulação desse campo com $k = 0.65$ e considerando as distribuições marginais univariadas, H e G , de $\mathbf{X}^{(0)}$ e $\mathbf{Y}^{(j)}$, $j \in \mathbb{N}$, respetivamente, definidas do seguinte modo,

$$H(x) = \exp(-x^{-1}), \quad x \in \mathbb{R}^+ \quad \text{e} \quad G(x) = \exp(-\exp^{-x}), \quad x \in \mathbb{R}.$$

Como consequência imediata do modo como \mathbf{X} foi definido, tem-se

$$X_{(i,j)} = k^j X_{(i,0)} \vee \bigvee_{t=1}^j k^{j-t+1} Y_{(i,t)}, \quad (i,j) \in \mathbb{N}^2, \quad 0 < k < 1. \quad (5.3)$$

Denotemos as distribuições de dimensão finita de $\mathbf{X}^{(0)}$ e do campo das inovações $\mathbf{Y} = \{Y_{(i,j)}\}_{(i,j) \in \mathbb{N}^2}$, formado pelas variáveis das sucessões $\mathbf{Y}^{(j)}$, respetivamente, por $H_{(i_1,0), \dots, (i_p,0)}(x_1, \dots, x_p)$ e $G_{(i_1, j_1), \dots, (i_p, j_p)}(x_1, \dots, x_p)$, com $\{i_1, \dots, i_p\} \subset \mathbb{N}$, $\{j_1, \dots, j_p\} \subset \mathbb{N}$, $(x_1, \dots, x_p) \in \mathbb{R}^p$, isto é,

$$P(X_{(i_1,0)} \leq x_1, \dots, X_{(i_p,0)} \leq x_p) = H_{(i_1,0), \dots, (i_p,0)}(x_1, \dots, x_p)$$

e

$$P(Y_{(i_1, j_1)} \leq x_1, \dots, Y_{(i_p, j_p)} \leq x_p) = G_{(i_1, j_1), \dots, (i_p, j_p)}(x_1, \dots, x_p).$$

Para quaisquer localizações $\mathbf{i}_1 = (i_1, j_1), \dots, \mathbf{i}_p = (i_p, j_p)$ em \mathbb{N}^2 , atendendo a (5.3), a função de distribuição conjunta de $X_{\mathbf{i}_1}, \dots, X_{\mathbf{i}_p}$ é dada por

$$\begin{aligned} H_{\mathbf{i}_1, \dots, \mathbf{i}_p}(x_1, \dots, x_p) &= P(X_{(i_1, j_1)} \leq x_1, \dots, X_{(i_p, j_p)} \leq x_p) \\ &= P\left(X_{(i_1, 0)} \leq \frac{x_1}{k^{j_1}}, \dots, X_{(i_p, 0)} \leq \frac{x_p}{k^{j_p}}\right) \times \\ &\quad \times P\left(\bigvee_{t=1}^{j_1} k^{j_1-t+1} Y_{(i_1, t)} \leq x_1, \dots, \bigvee_{t=1}^{j_p} k^{j_p-t+1} Y_{(i_p, t)} \leq x_p\right), \end{aligned}$$

ou, de forma equivalente, com base em (5.2),

$$\begin{aligned} H_{\mathbf{i}_1, \dots, \mathbf{i}_p}(x_1, \dots, x_p) &= P\left(X_{(i_1, j_1-1)} \leq \frac{x_1}{k}, \dots, X_{(i_p, j_p-1)} \leq \frac{x_p}{k}\right) \times \\ &\quad \times P\left(Y_{(i_1, j_1)} \leq \frac{x_1}{k}, \dots, Y_{(i_p, j_p)} \leq \frac{x_p}{k}\right) \\ &= H_{\mathbf{i}_1+(0, -1), \dots, \mathbf{i}_p+(0, -1)}\left(\frac{x_1}{k}, \dots, \frac{x_p}{k}\right) \times G_{\mathbf{i}_1, \dots, \mathbf{i}_p}\left(\frac{x_1}{k}, \dots, \frac{x_p}{k}\right). \end{aligned}$$

Tal como referimos anteriormente, se fixarmos $i_1 = \dots = i_p = i \geq 1$ obtém-se a sucessão Markovian estudada por Alpuim (1989), para a qual se sabe que

$$H_{(i, j_1), \dots, (i, j_p)}(x_1, \dots, x_p) = H\left(\bigwedge_{s=1}^p \frac{x_s}{k^{j_s}}\right) \prod_{t=1}^p \prod_{s=j(t-1)}^{j_t-1} G\left(\bigwedge_{m=t}^p \frac{x_m}{k^{j_m-s}}\right), \quad (5.4)$$

para $0 = j_0 < j_1 < \dots < j_p$. Este resultado permite definir a f.d. conjunta das variáveis $X_{(i, j_1)}, \dots, X_{(i, j_p)}$ desde que se conheçam as f.d.'s marginais univariadas H e G .

No que se segue, consideram-se as localizações $\mathbf{i}_1 = (i_1, j_1), \dots, \mathbf{i}_p = (i_p, j_p)$ em \mathbb{N}^2 , tais que $i_{m_1} \neq i_{m_2}$, $m_1, m_2 \in \{1, \dots, p\}$, porque a situação $i_{m_1} = i_{m_2}$ já foi estudada, como acabámos de referir.

Na próxima secção iremos averiguar quais são as condições necessárias e suficientes que devem ser verificadas pelo campo aleatório ARMAX, \mathbf{X} , para que se tenha a estacionariedade.

5.3 Estacionariedade

O próximo resultado evidencia a relação que deverá existir entre as distribuições de dimensão finita do campo ARMAX, \mathbf{X} , e do campo das inovações, \mathbf{Y} , para que se tenha a estacionariedade de \mathbf{X} .

Proposição 5.1 *O campo ARMAX, \mathbf{X} , definido em (5.2) é estacionário se e só se para quaisquer localizações $\mathbf{i}_1, \dots, \mathbf{i}_p \in \mathbb{N}^2$ e $(x_1, \dots, x_p) \in \mathbb{R}^p$,*

$$H_{\mathbf{i}_1, \dots, \mathbf{i}_p}(x_1, \dots, x_p) = H_{\mathbf{i}_1, \dots, \mathbf{i}_p} \left(\frac{x_1}{k}, \dots, \frac{x_p}{k} \right) \times G_{\mathbf{i}_1, \dots, \mathbf{i}_p} \left(\frac{x_1}{k}, \dots, \frac{x_p}{k} \right). \quad (5.5)$$

Demonstração: Considere-se que o campo aleatório \mathbf{X} é estacionário com distribuições de dimensão finita $H_{\mathbf{i}_1, \dots, \mathbf{i}_p}$, $\mathbf{i}_1, \dots, \mathbf{i}_p \in \mathbb{N}^2$. Então,

$$\begin{aligned} H_{\mathbf{i}_1+(0,n), \dots, \mathbf{i}_p+(0,n)}(x_1, \dots, x_p) &= P(X_{(i_1, j_1+n)} \leq x_1, \dots, X_{(i_p, j_p+n)} \leq x_p) \\ &= H_{\mathbf{i}_1+(0, n-1), \dots, \mathbf{i}_p+(0, n-1)} \left(\frac{x_1}{k}, \dots, \frac{x_p}{k} \right) \\ &\quad \times G_{(i_1, j_1), \dots, (i_p, j_p)} \left(\frac{x_1}{k}, \dots, \frac{x_p}{k} \right), \end{aligned}$$

e, pela estacionariedade de \mathbf{X} , obtém-se

$$\begin{aligned} H_{\mathbf{i}_1, \dots, \mathbf{i}_p}(x_1, \dots, x_p) &= \lim_{n \rightarrow +\infty} H_{\mathbf{i}_1+(0,n), \dots, \mathbf{i}_p+(0,n)}(x_1, \dots, x_p) \\ &= H_{\mathbf{i}_1, \dots, \mathbf{i}_p} \left(\frac{x_1}{k}, \dots, \frac{x_p}{k} \right) \times G_{\mathbf{i}_1, \dots, \mathbf{i}_p} \left(\frac{x_1}{k}, \dots, \frac{x_p}{k} \right). \end{aligned}$$

Reciprocamente, se (5.5) se verifica, então

$$\begin{aligned} H_{\mathbf{i}_1, \dots, \mathbf{i}_p}(x_1, \dots, x_p) &= H_{\mathbf{i}_1+(0,-1), \dots, \mathbf{i}_p+(0,-1)} \left(\frac{x_1}{k}, \dots, \frac{x_p}{k} \right) \times G_{\mathbf{i}_1, \dots, \mathbf{i}_p} \left(\frac{x_1}{k}, \dots, \frac{x_p}{k} \right) \\ &= H_{\mathbf{i}_1+(0,-1), \dots, \mathbf{i}_p+(0,-1)} \left(\frac{x_1}{k}, \dots, \frac{x_p}{k} \right) \times \\ &\quad \times \frac{H_{\mathbf{i}_1+(0,-1), \dots, \mathbf{i}_p+(0,-1)}(x_1, \dots, x_p)}{H_{\mathbf{i}_1+(0,-1), \dots, \mathbf{i}_p+(0,-1)} \left(\frac{x_1}{k}, \dots, \frac{x_p}{k} \right)} \\ &= H_{\mathbf{i}_1+(0,-1), \dots, \mathbf{i}_p+(0,-1)}(x_1, \dots, x_p), \quad (x_1, \dots, x_p) \in \mathbb{R}^p. \end{aligned}$$

Por outro lado, atendendo a que as sucessões $\mathbf{X}^{(0)}$ e $\mathbf{Y}^{(j)}$, $j \in \mathbb{N}$, são estacionárias, tem-se

$$H_{\mathbf{i}_1, \dots, \mathbf{i}_p}(x_1, \dots, x_p) = H_{\mathbf{i}_1+(n,0), \dots, \mathbf{i}_p+(n,0)}(x_1, \dots, x_p), \quad n \in \mathbb{N}, \quad (x_1, \dots, x_p) \in \mathbb{R}^p,$$

ficando provado que \mathbf{X} é estacionário. \square

Este resultado mostra que partindo da sucessão $\mathbf{X}^{(0)}$, com distribuições de dimensão finita $H_{(i_1,0),\dots,(i_p,0)}$, $i_1, \dots, i_p \in \mathbb{N}$, as distribuições de dimensão finita de cada $\mathbf{Y}^{(j)}$, $j \in \mathbb{N}$, têm de ser definidas por

$$G_{(i_1,j),\dots,(i_p,j)}(x_1, \dots, x_p) = \frac{H_{(i_1,0),\dots,(i_p,0)}(kx_1, \dots, kx_p)}{H_{(i_1,0),\dots,(i_p,0)}(x_1, \dots, x_p)}, \quad (x_1, \dots, x_p) \in \mathbb{R}^p. \quad (5.6)$$

Por outro lado, sendo $\mathbf{Y}^{(j)}$, $j \in \mathbb{N}$, sucessões independentes, para quaisquer $i_1, \dots, i_p, j_1, \dots, j_p \in \mathbb{N}$, a f.d. conjunta $G_{(i_1,j_1),\dots,(i_p,j_p)}(x_1, \dots, x_p)$ corresponderá ao produto de quocientes análogos a (5.6).

O estudo feito nos capítulos precedentes recaiu sobre campos aleatórios max-estáveis. Na próxima secção estabelecemos uma condição necessária e suficiente para que o campo ARMAX verifique a propriedade de max-estabilidade.

5.4 Max-estabilidade

Se as distribuições de dimensão finita das sucessões $\mathbf{Y}^{(j)}$, $j \geq 1$, associadas ao campo aleatório das inovações, \mathbf{Y} , são multivariadas de valores extremos então \mathbf{Y} é um campo aleatório max-estável.

O próximo resultado é uma condição necessária e suficiente para que o campo ARMAX, \mathbf{X} , seja max-estável.

Proposição 5.2 *O campo ARMAX, \mathbf{X} , definido em (5.2) é max-estável se e só se \mathbf{Y} é um campo aleatório max-estável.*

Demonstração: Para quaisquer localizações $\mathbf{i}_1 = (i_1, j_1), \dots, \mathbf{i}_p = (i_p, j_p) \in \mathbb{N}^2$ suponhamos que $G_{\mathbf{i}_1, \dots, \mathbf{i}_p}$ é uma distribuição MEV, ou seja, a sua função de dependência satisfaz a equação de estabilidade (1.9), isto é, para todo o $t \geq 1$

$$D_{G_{\mathbf{i}_1, \dots, \mathbf{i}_p}}^t(u_1^{1/t}, \dots, u_p^{1/t}) = D_{G_{\mathbf{i}_1, \dots, \mathbf{i}_p}}(u_1, \dots, u_p), \quad u_i \in [0, 1], \quad i = 1, \dots, p,$$

e as margens univariadas são do tipo extremo.

Além disso, assumamos, sem perda de generalidade, que a distribuição marginal univariada de \mathbf{X} é Fréchet unitária, $H(x) = H_{\mathbf{i}}(x) = \exp(-x^{-1})$, $x > 0$, $\mathbf{i} \in \mathbb{N}^2$. Assim, as distribuições marginais univariadas de \mathbf{Y} também são de Fréchet, sendo

$$G(x) = G_{\mathbf{i}}(x) = \frac{H_{\mathbf{i}}(kx)}{H_{\mathbf{i}}(x)} = \exp\left(-\left(\frac{kx}{1-k}\right)^{-1}\right) = H\left(\frac{kx}{1-k}\right), \quad x > 0, \quad \mathbf{i} \in \mathbb{N}^2.$$

A partir desta caracterização das distribuições univariadas de \mathbf{X} e \mathbf{Y} obtém-se

$$\begin{aligned}
 D_{H_{i_1, \dots, i_p}}^t(u_1^{\frac{1}{t}}, \dots, u_p^{\frac{1}{t}}) &= D_{H_{i_1, \dots, i_p}}^t(H^{\frac{1}{t}}(x_1), \dots, H^{\frac{1}{t}}(x_p)) \\
 &= D_{H_{i_1, \dots, i_p}}^t(H(tx_1), \dots, H(tx_p)) \\
 &= H_{i_1, \dots, i_p}^t(tx_1, \dots, tx_p) \\
 &= H_{i_1, \dots, i_p}^t\left(\frac{tx_1}{k}, \dots, \frac{tx_p}{k}\right) \times G_{i_1, \dots, i_p}^t\left(\frac{tx_1}{k}, \dots, \frac{tx_p}{k}\right) \\
 &= D_{H_{i_1, \dots, i_p}}^t\left(H\left(\frac{tx_1}{k}\right), \dots, H\left(\frac{tx_p}{k}\right)\right) \times \\
 &\quad \times D_{G_{i_1, \dots, i_p}}^t\left(G\left(\frac{tx_1}{k}\right), \dots, G\left(\frac{tx_p}{k}\right)\right) \\
 &= D_{H_{i_1, \dots, i_p}}^t(u_1^{\frac{k}{t}}, \dots, u_p^{\frac{k}{t}}) D_{G_{i_1, \dots, i_p}}^t(u_1^{\frac{1-k}{t}}, \dots, u_p^{\frac{1-k}{t}}) \\
 &= D_{H_{i_1, \dots, i_p}}^t(u_1^{\frac{k^2}{t}}, \dots, u_p^{\frac{k^2}{t}}) D_{G_{i_1, \dots, i_p}}^t(u_1^{k(1-k)}, \dots, u_p^{k(1-k)}) \times \\
 &\quad \times D_{G_{i_1, \dots, i_p}}^t(u_1^{1-k}, \dots, u_p^{1-k}) \\
 &= D_{H_{i_1, \dots, i_p}}^t(u_1^{\frac{k^n}{t}}, \dots, u_p^{\frac{k^n}{t}}) \prod_{l=0}^{n-1} D_{G_{i_1, \dots, i_p}}^{k^l}(u_1^{1-k}, \dots, u_p^{1-k})
 \end{aligned}$$

de onde se tem, para $n \rightarrow +\infty$,

$$\begin{aligned}
 \ln D_{H_{i_1, \dots, i_p}}^t(u_1^{1/t}, \dots, u_p^{1/t}) &= 0 + \ln D_{G_{i_1, \dots, i_p}}(u_1^{1-k}, \dots, u_p^{1-k}) \sum_{l=0}^{\infty} k^l \\
 &= \ln D_{G_{i_1, \dots, i_p}}^{\frac{1}{1-k}}(u_1^{1-k}, \dots, u_p^{1-k}).
 \end{aligned}$$

Assim,

$$D_{H_{i_1, \dots, i_p}}^t(u_1^{1/t}, \dots, u_p^{1/t}) = D_{G_{i_1, \dots, i_p}}(u_1, \dots, u_p), \quad t \geq 1,$$

e, consequentemente, tem-se

$$D_{H_{i_1, \dots, i_p}}^t(u_1^{1/t}, \dots, u_p^{1/t}) = D_{H_{i_1, \dots, i_p}}(u_1, \dots, u_p), \quad t \geq 1.$$

Suponhamos agora que H_{i_1, \dots, i_p} é MEV. Então,

$$\begin{aligned}
 D_{G_{i_1, \dots, i_p}}^t(u_1^{\frac{1}{t}}, \dots, u_p^{\frac{1}{t}}) &= D_{G_{i_1, \dots, i_p}}^t(G^{\frac{1}{t}}(x_1), \dots, G^{\frac{1}{t}}(x_p)) \\
 &= D_{G_{i_1, \dots, i_p}}^t(G(tx_1), \dots, G(tx_p)) \\
 &= \frac{D_{H_{i_1, \dots, i_p}}^t(H(ktx_1), \dots, H(ktx_p))}{D_{H_{i_1, \dots, i_p}}^t(H(tx_1), \dots, H(tx_p))} \\
 &= \frac{D_{H_{i_1, \dots, i_p}}^t(u_1^{\frac{1}{1-k} \frac{1}{t}}, \dots, u_p^{\frac{1}{1-k} \frac{1}{t}})}{D_{H_{i_1, \dots, i_p}}^t(u_1^{\frac{k}{1-k} \frac{1}{t}}, \dots, u_p^{\frac{k}{1-k} \frac{1}{t}})} \\
 &= \frac{D_{H_{i_1, \dots, i_p}}^t(u_1^{\frac{1}{1-k}}, \dots, u_p^{\frac{1}{1-k}})}{D_{H_{i_1, \dots, i_p}}^t(u_1^{\frac{k}{1-k}}, \dots, u_p^{\frac{k}{1-k}})} \\
 &= \frac{D_{H_{i_1, \dots, i_p}}^t(H(kx_1), \dots, H(kx_p))}{D_{H_{i_1, \dots, i_p}}^t(H(x_1), \dots, H(x_p))} \\
 &= G_{i_1, \dots, i_p}(x_1, \dots, x_p).
 \end{aligned}$$

□

Depois de estabelecidas condições para que o campo ARMAX, \mathbf{X} , seja estacionário e max-estável, segue-se a determinação de alguns coeficientes de dependência espacial.

5.5 Coeficientes de dependência

É possível relacionar os coeficientes extremais do campo ARMAX, \mathbf{X} , com os do campo das inovações, \mathbf{Y} , como se mostra no resultado que se apresenta de seguida.

Proposição 5.3 *Se \mathbf{X} e \mathbf{Y} são campos aleatórios estacionários e max-estáveis então os coeficientes extremais das suas distribuições de dimensão finita coincidem.*

Demonstração: Se \mathbf{X} (\mathbf{Y}) é um campo aleatório estacionário e max-estável então de (5.5) obtém-se

$$\begin{aligned}
 H^{\varepsilon^{\mathbf{X}}(\{\mathbf{i}_1, \dots, \mathbf{i}_p\})}(kx) &= H^{\varepsilon^{\mathbf{X}}(\{\mathbf{i}_1, \dots, \mathbf{i}_p\})}(x) G^{\varepsilon^{\mathbf{Y}}(\{\mathbf{i}_1, \dots, \mathbf{i}_p\})}(x) \Leftrightarrow \\
 \Leftrightarrow \varepsilon^{\mathbf{X}}(\{\mathbf{i}_1, \dots, \mathbf{i}_p\}) \ln H(kx) &= \varepsilon^{\mathbf{X}}(\{\mathbf{i}_1, \dots, \mathbf{i}_p\}) \ln H(x) + \varepsilon^{\mathbf{Y}}(\{\mathbf{i}_1, \dots, \mathbf{i}_p\}) \ln G(x) \Leftrightarrow \\
 \Leftrightarrow (\varepsilon^{\mathbf{X}}(\{\mathbf{i}_1, \dots, \mathbf{i}_p\}) - \varepsilon^{\mathbf{Y}}(\{\mathbf{i}_1, \dots, \mathbf{i}_p\})) \ln G(x) &= 0
 \end{aligned}$$

Assim, se G é uma distribuição não degenerada tem-se

$$\varepsilon^{\mathbf{X}}(\{\mathbf{i}_1, \dots, \mathbf{i}_p\}) = \varepsilon^{\mathbf{Y}}(\{\mathbf{i}_1, \dots, \mathbf{i}_p\}).$$

□

Note-se que o coeficiente de dependência $\varepsilon^{\mathbf{Y}}$ é a soma dos coeficientes de dependência das sucessões independentes $\mathbf{Y}^{(j)}$, $j \geq 1$.

Se se considerarem as localizações $\mathbf{i}_1, \dots, \mathbf{i}_p$ em \mathbb{N}^2 tais que

$$i_1 = \dots = i_p = i \geq 1 \quad \text{e} \quad 0 = j_0 < j_1 < \dots < j_p,$$

o cálculo do coeficiente extremal depende das f.d.'s de dimensão finita

$$H_{(i, j_1), \dots, (i, j_p)}(x, \dots, x), \quad x \in \mathbb{R},$$

caracterizadas em (5.4), e apresenta-se na proposição seguinte.

Proposição 5.4 *Seja \mathbf{X} o campo aleatório ARMAX tal como definido em (5.2), estacionário e max-estável. Então, para quaisquer $0 = j_0 < j_1 < \dots < j_p$ e $\mathbf{A} = \{(i, j_1), \dots, (i, j_p)\}$, obtém-se*

$$\varepsilon(\mathbf{A}) = k^{j_1} + \sum_{t=1}^p (1 - k^{j_t - j_{(t-1)}}).$$

Demonstração: Atendendo a (5.4), obtém-se

$$\begin{aligned}
 \varepsilon(\mathbf{A}) &= \frac{\ln \left(H \left(\frac{x}{k^{j_1}} \right) \right) + \sum_{t=1}^p \sum_{s=j_{(t-1)}}^{j_t-1} \ln H \left(\frac{x}{k^{j_t-s-1}} \right) - \ln H \left(\frac{x}{k^{j_t-s}} \right)}{\ln H(x)} \\
 &= k^{j_1} + \sum_{t=1}^p \sum_{s=j_{(t-1)}}^{j_t-1} \frac{\ln H^{k^{j_t-s-1}}(x) - \ln H^{k^{j_t-s}}(x)}{\ln H(x)}
 \end{aligned}$$

$$\begin{aligned}
 &= k^{j_1} + \sum_{t=1}^p \sum_{s=j_{(t-1)}}^{j_t-1} (k^{j_t-s-1} - k^{j_t-s}) \\
 &= k^{j_1} + \sum_{t=1}^p (1 - k^{j_t-j_{(t-1)}}).
 \end{aligned}$$

□

Os dois resultados anteriores serão fundamentais para avaliar a dependência extremal do campo aleatório \mathbf{X} através dos coeficientes introduzidos nos capítulos anteriores.

No que se segue, consideramos o campo ARMAX, \mathbf{X} , definido em (5.2), onde a função de dependência das distribuições de dimensão finita da sucessão $\mathbf{X}^{(0)}$ é a função Gumbel com parâmetro $\delta \geq 1$, isto é,

$$D_{H_{(i_1,0),\dots,(i_p,0)}}(u_1, \dots, u_p) = \exp \left(- \left(\sum_{i=1}^p (-\ln u_i)^\delta \right)^{1/\delta} \right), \quad u_i \in [0, 1]. \quad (5.7)$$

Nestas condições, e atendendo a (5.6), para cada $j \in \mathbb{N}$, a f.d. conjunta das variáveis $Y_{(i_1,j)}, \dots, Y_{(i_p,j)}$ da sucessão $\mathbf{Y}^{(j)}$, é dada por

$$\begin{aligned}
 G_{(i_1,j),\dots,(i_p,j)}(x_1, \dots, x_p) &= \frac{\exp \left(- \left(\sum_{j=1}^p \left(\frac{1}{kx_j} \right)^\delta \right)^{1/\delta} \right)}{\exp \left(- \left(\sum_{j=1}^p \left(\frac{1}{x_j} \right)^\delta \right)^{1/\delta} \right)} \\
 &= \exp \left(- \left(\sum_{j=1}^p \left(\frac{1}{kx_j} \right)^\delta \right)^{1/\delta} + \left(\sum_{j=1}^p \left(\frac{1}{x_j} \right)^\delta \right)^{1/\delta} \right) \\
 &= \exp \left((1 - k^{-1}) \left(\sum_{j=1}^p \left(\frac{1}{x_j} \right)^\delta \right)^{1/\delta} \right). \quad (5.8)
 \end{aligned}$$

Definidas as f.d.'s de dimensão finita da sucessão de partida $\mathbf{X}^{(0)}$ e da sucessão $Y^{(j)}, j \in \mathbb{N}$, temos os elementos necessários para avaliar a dependência extremal no campo aleatório ARMAX, \mathbf{X} , através dos coeficientes de dependência espacial introduzidos nos capítulos anteriores. Estes coeficientes abrangem diferentes vertentes da dependência extremal, como ilustramos no que segue. Começaremos por recorrer às matrizes de coeficientes de dependência na cauda superior (2.8), para quantificar a dependência extremal entre as variáveis localizadas em $\mathbf{i} = (i, j) \in \mathbb{N}^2$ e em

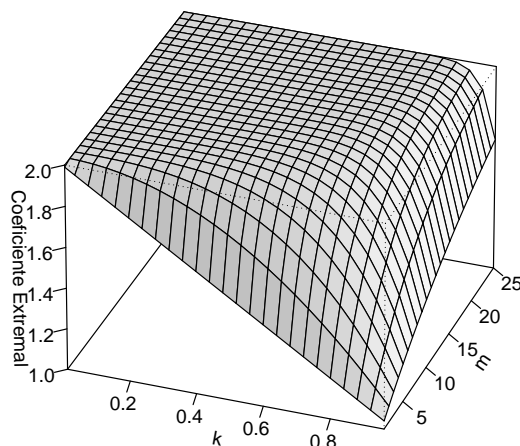


Figura 5.2: Valores de $\varepsilon(\mathbf{i}, T_{s_3}^m(\mathbf{i}))$ para $m \in \{1, \dots, 25\}$ e $k \in \{0.01, 0.04m : m \in \{1, \dots, 24\}\}$.

localizações vizinhas de \mathbf{i} . Com o madograma generalizado iremos avaliar a dependência entre os máximos das variáveis de \mathbf{X} localizadas em duas regiões disjuntas. Seguir-se-á o cálculo do índice de contágio, para quantificarmos a propagação de um evento extremo numa localização \mathbf{i} a uma região vizinha e, por fim, a estabilidade das variáveis de \mathbf{X} numa região, face a $X_{\mathbf{i}}$, será traduzida pelo índice de estabilidade.

5.5.1 Matrizes de coeficientes de dependência na cauda superior

Pela Proposição 5.4, para qualquer $\mathbf{i} = (i, j) \in \mathbb{N}^2$, pode concluir-se que

$$\varepsilon(\mathbf{i}, T_{s_3}^m(\mathbf{i})) = \varepsilon(\{(i, j), (i, j + m)\}) = 2 - k^m, \quad m \geq 1. \quad (5.9)$$

A variação de $\varepsilon(\mathbf{i}, T_{s_3}^m(\mathbf{i}))$ (Figura 5.2) depende do parâmetro k do campo aleatório \mathbf{X} e do passo m , que se toma na definição dos vizinhos de \mathbf{i} segundo a direção s_3 . O valor deste coeficiente tende para 2 (independência entre os acontecimentos $\{X_{\mathbf{i}} \leq x\}$ e $\{X_{T_{s_3}^m(\mathbf{i})} \leq x\}$) quando $m \rightarrow +\infty$.

Aplicando a relação (1.16), obtém-se

$$\lambda(T_{s_3}^m(\mathbf{i}), \mathbf{i}) = k^m.$$

Embora o coeficiente extremal associado às variáveis $X_{\mathbf{i}}$ e $X_{T_{s_3}^m(\mathbf{i})}$, dado em (5.9) não dependa das f.d.'s de dimensão finita de $\mathbf{X}^{(0)}$ e de $\mathbf{Y}^{(j)}$, $j \in \mathbb{N}$, tal não se verifica quando consideramos o coeficiente extremal associado às v.a.'s indexadas na

3	3	3	3	3	3	3
3	2	2	2	2	2	3
3	2	1	1	1	2	3
3	2	1	1	1	2	3
3	2	1	1	1	2	3
3	2	2	2	2	2	3
3	3	3	3	3	3	3

Figura 5.3: Representação, numa escala de cinzas, dos valores da matriz $\Lambda^3(\mathbf{i})$.

região $\{(i_1, 0), \dots, (i_p, 0)\}$. De facto,

$$\varepsilon(\{(i_1, 0), \dots, (i_p, 0)\}) = p^{1/\delta}, \quad p > 1, \delta \geq 1, \quad (5.10)$$

para além de depender do número de variáveis consideradas depende também do parâmetro δ associado à função de dependência de $H_{(i_1, 0), \dots, (i_p, 0)}$, definida em (5.7).

Considerando (5.9) e (5.10) obtém-se, para $m \geq 1$, a seguinte matriz de coeficientes de dependência na cauda,

$$\Lambda(T_s^m(\mathbf{i}), \mathbf{i}) = \begin{bmatrix} 0 & k^m & 0 \\ 2 - 2^{1/\delta} & 1 & 2 - 2^{1/\delta} \\ 0 & k^m & 0 \end{bmatrix}. \quad (5.11)$$

O cálculo dos valores da matriz de dependência $\Lambda^3(\mathbf{i})$, definida em (2.10), é imediato e encontram-se representados na Figura 5.3 através de uma escala de cinzas, onde o branco indica independência, o preto dependência total e a numeração corresponde ao passo m que se está a considerar. Observando esta representação pode dizer-se que existe independência entre os acontecimentos $\{X_i \leq x\}$ e $\{X_j \leq x\}$ desde que \mathbf{j} seja um vizinho de \mathbf{i} que se alcança aplicando uma qualquer composição de direções, definidas em (2.6), diferente de $s_j^m(\mathbf{i})$, $j \in \{1, 3, 5, 7\}$, o que já era visível na matriz (5.11).

5.5.2 Madograma generalizado

Vamos agora considerar o madograma generalizado, apresentado na Secção 3.2, para avaliar a dependência entre o máximo das variáveis de \mathbf{X} indexadas em duas regiões disjuntas de localizações.

Considerando, por exemplo, os seguintes conjuntos de localizações

$$\mathbf{I} = \{(2, 1), (2, 2)\} \quad \text{e} \quad \mathbf{J} = \{(3, 3), (3, 4)\},$$

tem-se

$$V_{\mathbf{I},\mathbf{J}}(\alpha, \alpha, \beta, \beta) = -\ln(H_{\mathbf{I},\mathbf{J}}(\alpha, \alpha, \beta, \beta)).$$

Aplicando sucessivamente (5.2) à f.d. $H_{\mathbf{I},\mathbf{J}}(\alpha, \alpha, \beta, \beta)$ obtém-se

$$\begin{aligned} H_{\mathbf{I},\mathbf{J}}(\alpha, \alpha, \beta, \beta) &= H_{(2,0),(3,0)}\left(\frac{\alpha}{k}, \frac{\beta}{k^3}\right) \times G_{(2,1),(3,1)}\left(\frac{\alpha}{k}, \frac{\beta}{k^3}\right) \times \\ &\quad \times G_{(2,2),(3,2)}\left(\frac{\alpha}{k}, \frac{\beta}{k^2}\right) \times G_{(3,3)}\left(\frac{\beta}{k}\right) \times G_{(3,4)}\left(\frac{\beta}{k}\right), \end{aligned}$$

e dado que a f.d. univariada de $\mathbf{X}^{(0)}$ é Fréchet unitária, $H_{(i,0)}(x) = \exp(-x^{-1})$, $i \in \mathbb{N}$, e $G_{(i_1,j),\dots,(i_p,j)}(x_1, \dots, x_p)$ é definida por (5.8), vem

$$\begin{aligned} V_{\mathbf{I},\mathbf{J}}(\alpha, \alpha, \beta, \beta) &= \left[\left(\frac{k}{\alpha}\right)^\delta + \left(\frac{k^3}{\beta}\right)^\delta \right]^{1/\delta} - \\ &\quad - (1 - k^{-1}) \left[\left(\frac{k}{\alpha}\right)^\delta + \left(\frac{k^3}{\beta}\right)^\delta \right]^{1/\delta} - \\ &\quad - (1 - k^{-1}) \left[\left(\frac{k}{\alpha}\right)^\delta + \left(\frac{k^2}{\beta}\right)^\delta \right]^{1/\delta} - \\ &\quad - 2(1 - k^{-1}) \frac{k}{\beta}. \end{aligned} \tag{5.12}$$

De forma análoga obtém-se os seguintes valores da função expoente

$$V_{\mathbf{I}}(1, 1) = \varepsilon(\mathbf{I}) = -\ln(H_{\mathbf{I}}(1, 1)) = 2 - k = V_{\mathbf{J}}(1, 1) = \varepsilon(\mathbf{J}) = -\ln(H_{\mathbf{J}}(1, 1)). \tag{5.13}$$

Atendendo a (3.3), (5.12) e (5.13), o madograma generalizado associado às regiões \mathbf{I} e \mathbf{J} vem definido por

$$\begin{aligned} \nu^{\alpha,\beta}(\mathbf{I}, \mathbf{J}) &= \frac{(k^{-1}) \left[\left(\frac{k}{\alpha}\right)^\delta + \left(\frac{k^3}{\beta}\right)^\delta \right]^{1/\delta} - (1 - k^{-1}) \left[\left(\frac{k}{\alpha}\right)^\delta + \left(\frac{k^2}{\beta}\right)^\delta \right]^{1/\delta} - \frac{2k}{\beta} + \frac{2}{\beta}}{1 + k^{-1} \left[\left(\frac{k}{\alpha}\right)^\delta + \left(\frac{k^3}{\beta}\right)^\delta \right]^{1/\delta} - (1 - k^{-1}) \left[\left(\frac{k}{\alpha}\right)^\delta + \left(\frac{k^2}{\beta}\right)^\delta \right]^{1/\delta} - \frac{2k}{\beta} + \frac{2}{\beta}} - \\ &\quad - \frac{1}{2} \left(\frac{2 - k}{\alpha + 2 - k} + \frac{2 - k}{\beta + 2 - k} \right). \end{aligned} \tag{5.14}$$

O coeficiente fica dependente de k e δ , os parâmetros do campo ARMAX e da

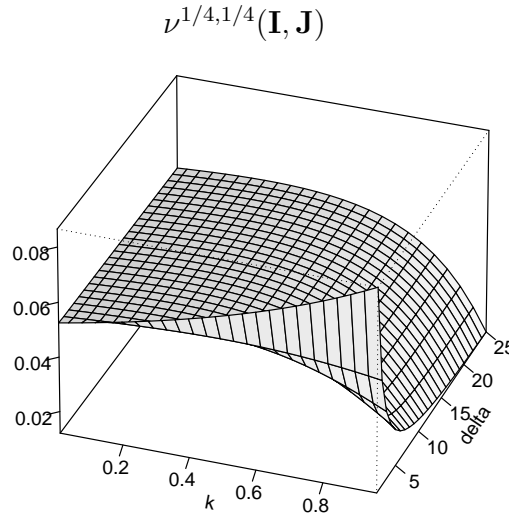


Figura 5.4: Representação do madograma generalizado (5.14) para $\delta \in \{1, \dots, 25\}$ e $k \in \{0.01, 0.04m : m \in \{1, \dots, 24\}\}$.

função de dependência Gumbel, respetivamente. Atribuindo igual peso a todas as variáveis localizadas em \mathbf{I} e \mathbf{J} , ou seja, tomando $\alpha = \beta = 1/4$, e considerando $\delta \in \{1, \dots, 25\}$ e $k \in \{0.01, 0.04m : m \in \{1, \dots, 24\}\}$ obtiveram-se os valores do madograma generalizado (5.14), representados na Figura 5.4. Observa-se que os valores mais elevados de k e δ conduzem aos valores mais baixos do madograma generalizado e, como tal, a uma maior dependência entre os máximos das variáveis de \mathbf{X} localizadas nas regiões \mathbf{I} e \mathbf{J} . Por outro lado, quando $\delta = 1$ as variáveis da sucessão $\mathbf{X}^{(0)}$ são independentes e os valores do madograma generalizado aumentam com o aumento de k , o que traduz uma menor dependência entre os referidos máximos.

5.5.3 Índices de contágio e estabilidade

Nesta secção vamos aplicar os índices de contágio e de estabilidade, introduzidos no Capítulo 4, para avaliar, respetivamente, a propagação de um acontecimento extremo numa localização \mathbf{i} a uma região vizinha e a estabilidade das variáveis de \mathbf{X} , localizadas numa dada região, face à localização \mathbf{i} .

Considerando a região $\mathbf{A} = \{(2, 4), (3, 4), (4, 4), (5, 4)\}$, a localização $\mathbf{i} = (3, 3)$ e atendendo a que pelas Proposições 5.3 e 5.4 se tem

$$\varepsilon(\{(3, 3), (2, 4)\}) = \varepsilon(\{(3, 3), (4, 4)\}) = \varepsilon(\{(3, 3), (5, 4)\}) = 2,$$

e

$$\varepsilon(\{(3, 3), (3, 4)\}) = 2 - k$$

o índice de contágio da localização \mathbf{i} para a região \mathbf{A} é dado por

$$CI(\mathbf{A}, \mathbf{i}) = 2|\mathbf{A}| - \sum_{\mathbf{j} \in \mathbf{A}} \varepsilon(\{\mathbf{i}, \mathbf{j}\}) = k.$$

A propagação da ocorrência de uma excedência de um nível elevado na localização \mathbf{i} à região \mathbf{A} , depende apenas do parâmetro $0 < k < 1$, o que mostra que esse acontecimento terá pouco impacto na região \mathbf{A} .

Vamos quantificar a estabilidade das variáveis de \mathbf{X} indexadas em \mathbf{A} , relativamente à localização \mathbf{i} , através do índice de estabilidade apresentado em (4.4). Obtém-se

$$SI(\mathbf{A}, \mathbf{i}) = \frac{|\mathbf{A}| - CI(\mathbf{A}, \mathbf{i})}{\varepsilon(\{\mathbf{i}\} \cup \mathbf{A}) - 1} = \frac{4 - k}{(3k^\delta + 1)^{1/\delta} - (k - 1)4^{1/\delta} - 1}, \quad (5.15)$$

uma vez que

$$\begin{aligned} \varepsilon(\{\mathbf{i}\} \cup \mathbf{A}) &= \ln \left(H_{(2,0),(3,0),(4,0),(5,0)} \left(\frac{x}{k^4}, \frac{x}{k^3}, \frac{x}{k^4}, \frac{x}{k^4} \right) \right) + \\ &+ \ln \left(G_{(2,1),(3,1),(4,1),(5,1)} \left(\frac{x}{k^4}, \frac{x}{k^3}, \frac{x}{k^4}, \frac{x}{k^4} \right) \right) + \\ &+ \ln \left(G_{(2,2),(3,2),(4,2),(5,2)} \left(\frac{x}{k^3}, \frac{x}{k^2}, \frac{x}{k^3}, \frac{x}{k^3} \right) \right) + \\ &+ \ln \left(G_{(2,3),(3,3),(4,3),(5,3)} \left(\frac{x}{k^2}, \frac{x}{k}, \frac{x}{k^2}, \frac{x}{k^2} \right) \right) + \\ &+ \ln \left(G_{(2,4),(3,4),(4,4),(5,4)} \left(\frac{x}{k}, \frac{x}{k}, \frac{x}{k}, \frac{x}{k} \right) \right) \\ &= (3k^\delta + 1)^{1/\delta} - (k - 1)4^{1/\delta}. \end{aligned} \quad (5.16)$$

Tal como no madograma generalizado, também neste caso o coeficiente $SI(\mathbf{A}, \mathbf{i})$ depende de k e δ , variando entre $\frac{4-k}{4}$ e $4 - k$.

Quando $\delta = 1$, situação de independência na sucessão $\mathbf{X}^{(0)}$, obtém-se $SI(\mathbf{A}, \mathbf{i}) = 1$, qualquer que seja o valor de k . Neste caso, a variação do campo ARMAX \mathbf{X} na região \mathbf{A} relativamente à localização \mathbf{i} é suave.

Observando os valores do coeficiente extremal representados na Figura 5.5 constatamos que são mais elevados quando $\delta = 1$, tendo-se neste caso $\varepsilon(\{\mathbf{i}\} \cup \mathbf{A}) = 5 - k$, sendo a dependência extremal entre as variáveis de \mathbf{X} localizadas em $\{\mathbf{i}\} \cup \mathbf{A}$ tanto menor quanto menor for o valor de k .

Quando $\delta \rightarrow +\infty$, situação de total dependência na sucessão $\mathbf{X}^{(0)}$, obtém-se $SI(\mathbf{A}, \mathbf{i}) = \frac{4-k}{1-k}$, sendo a variação do campo ARMAX \mathbf{X} na região \mathbf{A} face à localização \mathbf{i} dependente do parâmetro k . Tem-se $SI(\mathbf{A}, \mathbf{i})$ tanto maior quanto maior for o valor

5.5. COEFICIENTES DE DEPENDÊNCIA

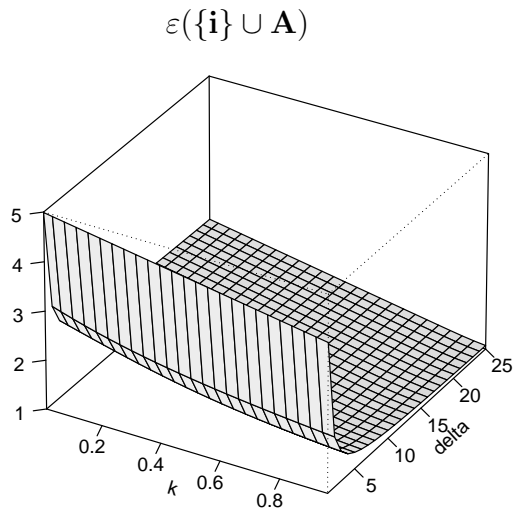


Figura 5.5: Representação do coeficiente extremal (5.16) para $\delta \in \{1, \dots, 25\}$ e $k \in \{0.01, 0.04m : m \in \{1, \dots, 24\}\}$.

de $0 < k < 1$.

Conclusões

O estudo que realizámos sobre a dependência extremal de campos aleatórios max-estáveis contribuiu com novos resultados teóricos que aplicámos em conjuntos de dados simulados e reais, permitindo mostrar o seu interesse prático.

Partindo da tradicional abordagem bivariada à dependência extremal em campos aleatórios max-estáveis, procurou-se fazer uma abordagem multivariada e multidirecional através de matrizes de coeficientes de dependência. Desta forma, centrámos o estudo na avaliação da dependência extremal entre variáveis indexadas em conjuntos de localizações, disjuntos e vizinhos segundo diferentes direções. O estudo prosseguiu no sentido de se definirem coeficientes que quantificassem a dependência extremal em situações específicas, com origem em problemas da vida real. Um dos coeficientes propostos foi o madograma generalizado cuja mais valia reside na quantificação da dependência entre os máximos em dois conjuntos de variáveis, localizadas em regiões disjuntas. Dado o interesse prático da avaliação da propagação de um evento extremo numa localização a uma região vizinha, introduziu-se o índice de contágio. Por outro lado, para quantificar a estabilidade de um campo aleatório max-estável sobre uma região, face a uma dada localização, propusemos o índice de estabilidade.

Para cada um dos coeficientes de dependência espacial propostos apresentámos estimadores não paramétricos e provámos as propriedades de consistência e normalidade assintótica. Em consequência das boas propriedades destes estimadores, aplicámo-los a amostras de dados obtidos por simulação e dados reais obtidos no SNIRH. No estudo de simulação utilizámos um campo aleatório M4 que definimos com base nos processos M4. Esta componente prática do trabalho exigiu a implementação de rotinas desenvolvidas no programa *R*. O facto deste programa ser de acesso livre, possibilita a divulgação e utilização do nosso trabalho pela comunidade científica.

Refira-se que a obtenção de amostras reais revelou-se uma tarefa difícil, dada a escassez de dados relativos a variáveis meteorológicas em Portugal. Tal facto limitou-nos aos dados de acesso livre disponíveis no SNIRH.

Com base no modelo introduzido por Haslett (1979) propôs-se um campo aleatório autoregressivo de máximos que denominámos de ARMAX. Neste modelo estu-

CONCLUSÕES

damos a estacionariedade, a max-estabilidade e a dependência extremal através das matrizes dos coeficientes de dependência multivariada e dos coeficientes introduzidos.

Identificámos algumas questões em aberto que poderão ser objeto de trabalho futuro. Exemplo disso é a possibilidade de definir novos estimadores para os coeficientes propostos e estudar as suas propriedades. Também teria interesse realizar estudos de simulação com outros campos aleatórios max-estáveis, apesar de termos obtido bons resultados com o estudo realizado sobre o campo M4. No âmbito de simulações com amostras de dados reais, seria desejável ter disponível dados de outras áreas que permitissem a aplicação dos coeficientes introduzidos.

Bibliografia

- Abrahamsen, P. (1997). *A review of gaussian random fields and correlation functions* (Tech. Rep.). Second Edition, Norwegian Computing Center.
- Adler, R. J. (1981). *The geometry of random fields*. John Wiley.
- Alpuim, M. T. (1989). An extremal markovian sequence. *J. Appl. Prob.*(26), 219-232.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., Waal, D. e Ferro, C. (2004). *Statistics of extremes: Theory and applications*. Wiley.
- Bel, L., Bacro, J. N. e Lantuéjoul, C. (2008). Assessing extremal dependence of environmental spatial fields. *Environmetrics*, 19, 163-182.
- Brown, B. e Resnick, S. (1977). Extreme values of independent stochastic processes. *J. Appl. Prob.*, 14, 732-739.
- Buishand, T. A. (1984). Bivariate extreme-value data and the station-year method. *J. Hydrol.*, 69, 77-95.
- Buishand, T. A., de Haan, L. e Zhou, C. (2008). On spatial extremes: With application to a rainfall problem. *Annals of Applied Statistics*, 2, 624-642.
- Coles, S. G. (1993). Regional modelling of extreme storms via max-stable processes. *J. R. Statist. Soc., Ser. B* 55, 797-816.
- Coles, S. G. (2001). *An introduction to statistical modeling of extreme values*. Springer-Verlag, London.
- Coles, S. G. e Tawn, J. A. (1996). Modelling extremes of the areal rainfall process. *J. R. Statist. Soc., Ser. B* 58, 329-347.
- Cooley, D. (2005). *Statistical analysis of extremes motivated by weather and climate studies: Applied and theoretical advances*. Tese de Doutorado, University of Colorado.
- Cooley, D., Naveau, P. e Poncet, P. (2006). Variograms for spatial max-stable random fields. *Lecture Notes in Statistics*, 187, 373-390.
- Daley, D. J. e Haslett, J. (1982). A thermal energy storage process with controlled input. *Adv. Appl. Probability*, 14, 257-271.
- Davison, A. C., Padon, S. A. e Ribatet, M. (2012). Statistical modelling of spatial extremes. *Statist. Sci.*, 27(2), 161-186.
- Deheuvels, P. (1978). Caractérisation complète des lois extrêmes multivariées et de la convergence aux types extrêmes. *Publ. Int. Statist. Unjiv. Paris*, 23, 1-36.
- Einmahl, J., Krajina, A. e Segers, J. (2012). An m-estimator for tail dependence in arbitrary dimensions. *arXiv*. Disponível em <http://arxiv.org/pdf/1112.0905.pdf>
- Einmahl, J., Li, J. e Liu, R. (2006). Extreme value theory approach to simultaneous monitoring and thresholding of multiple risk indicators. *Center Discussion*

- Paper (Int.rep. 2006-104) Econometrics.*
- Embrechts, P., Klüppelberg, C. e Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer: Berlin.
- Falk, M. e Tichy, D. (2011). Asymptotic conditional distribution of exceedance counts: fragility index with different margins. *Annals of the Institute of Statistical Mathematics*, 64, 1071-1085. Disponível em <http://dx.doi.org/10.1007/s10463-011-0348-3>
- Falk, M. e Tichy, D. (2012). Asymptotic conditional distribution of exceedance counts. *Adv. in Appl. Probab.*, 44 (1), 270-291.
- Fermanian, J. D., Radulovic, D. e Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10, 847-860.
- Ferreira, H. (2011). Dependence between two multivariate extremes. *Statis. & Probab. Letters*, 81 (5), 586-591.
- Ferreira, H. e Ferreira, M. (2012a). Fragility index of block tailed vectors. *J. Stat. Plan. Infer.*, 142, 1837-1848.
- Ferreira, H. e Ferreira, M. (2012b). On extremal dependence of block vectors. *Kybernetika*, 48, 988-1006.
- Ferreira, H. e Pereira, L. (2008). How to compute the extremal index of stationary random fields. *Statis. & Probab. Letters*, 78, 1301-1304.
- Ferreira, H. e Pereira, L. (2012). Point processes of exceedances by random fields. *J. Stat. Plan. Infer.*, 142, 773-779.
- Fisher, R. e Tippett, L. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proc. Cambridge Philos.*, 24, 180-190.
- Fonseca, C., Ferreira, H., Pereira, L. e Martins, A. (2012). Stability and contagion measures for spatial extreme value analyses. *Proceedings of the International Workshop on Spatio-Temporal Modelling (METMA VI)*.
- Fonseca, C., Martins, A., Pereira, L. e Ferreira, H. (2009). Matrices of local dependence between spatial extreme events. *Notas e Comunicações do CEAUL 14/09*. Disponível em <http://www.ceaul.fc.ul.pt/notas.html?ano=2009>
- Fonseca, C., Martins, A., Pereira, L. e Ferreira, H. (2010). Dependence matrices for spatial extreme events applied to a space-time armax storage model. *Proceedings of the International Workshop on Spatio-Temporal Modelling (METMA V)*, 160-163.
- Fonseca, C., Pereira, L., Ferreira, H. e Martins, A. (2011). Generalized madogram and pairwise dependence of maxima over two disjoint regions of a random field. *Proceedings of the 58th World Statistics Congress*. Disponível em <http://isi2011.congressplanner.eu/>
- Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Ann. de la Soc. Polonaise de Math.*, 6, 93.

- Galambos, J. (1978). *The asymptotic theory of extreme order statistics*. Wiley, New York.
- Geluk, J. L., de Haan, L. e Vries, C. G. de. (2007). *Weak and strong financial fragility* (Tech. Rep.). Tinbergen Institute Discussion Paper, TI 2007-023/2.
- Gilat, D. e Hill, T. (1992). One-sided refinements of the strong law of large numbers and the glivenko-cantelli theorem. *Ann. Probab.*, *20*, 1213-1221.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.*, *44*, 423-453.
- Greenwood, P. E. e Hooghiemstra, G. (1988). An extreme-type limit law for a storage process. *Math. Oper. Research*, *13*(2), 232-242.
- de Haan, L. (1970). *On regular variation and its application to the weak convergence of sample extremes*. Tese de Doutorado, Math. Centre Tracts 32, Mathematisch Centrum, Amsterdam.
- de Haan, L. e Ferreira, A. (2006). *Extreme value theory: An introduction*. Springer, New York.
- de Haan, L. e Lin, T. (1991). On convergence toward an extreme value distribution in $c[0,1]$. *Ann. Probab.*, *29*, 467-483.
- de Haan, L. e Pickands, J. (1986). Stationary min-stable stochastic processes. *Probab. Th. Rel. Fields*, *72*, 477-492.
- Haslett, J. (1979). A diffusion model for the storage of solar thermal energy. *J. Op. Res. Soc.*, *30*(5), 433-438.
- Hsing, T., Hüsler, J. e Leadbetter, M. (1988). On the exceedance point process for a stationary sequence. *Prob. Th. Rel. Fields*, *78*, 97-112.
- Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman & Hall, London.
- Kabluchko, Z. (2009). Extremes of space-time gaussian processes. *Stoch. Process and Their Applications*, *119*(11), 3962-3980.
- Krajina, A. (2010). *An m -estimator of multivariate tail dependence*. Tese de Doutorado, Tilburg University.
- Leadbetter, M. (1974). On extreme values in stationary sequences. *Z. wahrsc. verw. Gebiete*, *28*, 289-303.
- Leadbetter, M., Lindgren, G. e Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer-Verlag, New York.
- Leadbetter, M. e Nandagopalan, S. (1989). On exceedance point process for stationary sequences under mild oscillation restrictions. Em J. H. e D. Reiss (Ed.), *Extreme value theory* (p. 69-80). Springer-Verlag, New York.
- Li, H. (2009). Orthant tail dependence of multivariate extreme value distributions. *Journal of Multivariate Analysis*, *100*, 243-256.
- Martins, A. e Ferreira, H. (2005). The multivariate extremal index and the de-

Bibliografia

- pendence structure of a multivariate extreme value distribution . *Test*, 14(2), 433-448.
- von Mises, R. (1954). La distribution de la plus grande de n valeurs. *Selected papers II, Amer. Math. Soc.*, 271-294.
- Morales, F. C. (2005). *Estimation of max-stable processes using monte carlo methods with applications to financial risk assessment*. Tese de Doutorado, University of North Caroline.
- Nandagopalan, S. (1990). *Multivariate extremes and estimation of the extremal index*. Tese de Doutorado, University of North Caroline, Chapel Hill.
- Naveau, P., Guillou, A., Cooley, D. e Diebolt, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika*, 96, 1-17.
- Nelson, R. B. (2006). *An introduction to copulas*. Springer, New York.
- de Oliveira, J. T. (1962/63). Structure theory of bivariate extremes, extensions. *Est. Mat., Est. and Econ.*, 7, 165-195.
- Pereira, L. (2009). The asymptotic location of the maximum of a stationary random field. *Statis. & Probab. Letters*, 79, 2166-2169.
- Pereira, L. (2010). On the extremal behavior of a nonstationary normal random field. *J. Stat. Plan. Infer.*, 140, 3567-3576.
- Pereira, L. e Ferreira, H. (2006). Limiting crossing probabilities of random fields. *J. Appl. Prob.*, 3, 884-891.
- Pickands, J. (1981). Multivariate extreme value distributions. *Bulletin of the International Statistical Institute*, 49, 859-878.
- Piterbarg, V. I. (1996). Asymptotic methods in theory of gaussian processes and fields. *Trans. of Math. Monographs*, 148, American Mathematical Society.
- Resnick, S. (1987). *Extreme values, regular variation and point processes*. Springer-Verlag, New York.
- Resnick, S. (2006). *Heavy-tail phenomena: Probabilistic and statistical modeling*. Springer, New York.
- Ribatet, M. (2009). Spatial extremes: An r package for modeling spatial extremes. Em *Curso: Max-stable processes , theory and practice, ceaul, fcul, lisboa, 14 de outubro*.
- Sang, H. (2008). *Extreme value modeling for space-time data with metereological applications*. Tese de Doutorado, Duke University.
- Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1), 33-44.
- Schlather, M. (2012). *Simulation and analysis of random fields*. Disponível em <http://cran.r-project.org/web/packages/RandomFields/RandomFields.pdf>
- Schlather, M. e Tawn, J. (2003). A dependence measure for multivariate and spatial

- extremes: Properties and inference. *Biometrika*, *90*(1), 139-156.
- Schmidt, R. (2002). Tail dependence for elliptically countered distributions. *Math. Methods Oper. Res.*, *55*, 301-327.
- Sibuya, M. (1960). Bivariate extreme. *Ann. Inst. Statist. Math.*, *8*, 195-210.
- Smith, R. L. (1990). *Max-stable processes and spatial extremes* (Tech. Rep.). University of Surrey.
- Smith, R. L. e Weissman, I. (1996). *Characterization and estimation of the multivariate extremal index* (Tech. Rep.). Department of Statistics, University of North Carolina. Disponível em <http://www.stat.unc.edu/postscript/rs/extremal.pdf>
- Suveges, M. (2009). Modelling clusters of extremes events based on the m4 process. Em *Workshop on high-dimensional extremes*. Disponível em <http://extremes.epfl.ch/files/content/sites/extremes/files/users/111184/public/WorkshopSept09/Suveges.pdf>
- Trenberth, K. E. (1998). Atmospheric moisture residence times and cycling implications for rainfall rates and climate change. *Climate Change*, *39*, 667-694.
- Trenberth, K. E. (1999). Conceptual framework for changes of extremes of the hydrological cycle with climate change. *Climate Change*, *42*, 327-339.
- Vannitsem, S. e Naveau, P. (2007). Spatial dependences among precipitation maxima over Belgium. *Nonlin. Processes Geophys.*, *14*, 621-630.
- Viseu, C., Pereira, L., Martins, A. P. e Ferreira, H. (2012). On the multivariate upcrossings index. *Aceite para publicação em Communications in Statistics - Theory and Methods*.
- Zhang, Z. (2009). On approximating max-stable processes and constructing extremal copula function. *Stat. Inf. for Stoch. Proc.*, *12*, 89-114.
- Zhang, Z. e Smith, R. L. (2004). The behavior of multivariate maxima of moving maxima processes. *J. Appl. Probab.*, *41*, 1113-1123.
- Zhang, Z. e Smith, R. L. (2010). On the estimation and application of max-stable processes. *J. Stat. Plan. Infer.*, *140*, 1135-1153.