



Madogramas em campos aleatórios max-estáveis — Teoria e Aplicação

Versão final após defesa

Íris Almeida Rocha

Dissertação para obtenção do Grau de Mestre em
Matemática e Aplicações
(2º ciclo de estudos)

Orientadora: Prof. Doutora Luísa Maria Jota Pereira Amaral

Agosto de 2025

Declaração de Integridade

Eu, Íris Almeida Rocha, que abaixo assino, estudante com o número de inscrição M12466 do 2º ciclo de estudos em Matemática e Aplicações da Faculdade de Ciências, declaro ter desenvolvido o presente trabalho e elaborado o presente texto em total consonância com o **Código de Integridades da Universidade da Beira Interior**.

Mais concretamente afirmo não ter incorrido em qualquer das variedades de Fraude Académica, e que aqui declaro conhecer, que em particular atendi à exigida referenciação de frases, extratos, imagens e outras formas de trabalho intelectual, e assumindo assim na íntegra as responsabilidades da autoria.

Universidade da Beira Interior, Covilhã, 08/08/2025.

Íris Almeida Rocha

Agradecimentos

A escrita desta dissertação foi uma caminhada longa e íngreme, mas profundamente libertadora. Só posso expressar a minha mais sincera gratidão a todos aqueles que nunca me largaram a mão e tornaram este momento possível. À minha orientadora, a Professora Doutora Luísa Amaral, pelos inúmeros esclarecimentos, pelas correções e por todas as aprendizagens que me proporcionou. Por me ter ensinado a ser autónoma! A todos os professores que, de alguma forma, marcaram o meu percurso académico e me transmitiram ensinamentos que me acompanharão para sempre. Um agradecimento especial à Professora Doutora Ana Paula Martins, que me apresentou à Teoria de Valores Extremos e com quem eu aprendi tanto! À minha amiga Natalie, pela sua fé inabalável nas minhas capacidades, pela sua presença constante e pelas conversas profundas e enternecedoras. À Liliana, por me despertar, por me mostrar que tudo estava dentro de mim e que, afinal, o que parecia uma montanha não passava de uma ervilha. À Maria, à Ana Maria e à Valentina, pelas conversas, pelos passeios e pelos bolos que me aqueceram a alma e me mantiveram à tona. Aos meus pais, por serem o meu porto seguro. À minha mãe, por me ouvir sempre e por fazer o possível e impossível por mim. Ao meu pai, por me incentivar constantemente a ser melhor e a saber mais. A mim mesma. A Deus. Acima de tudo, a Deus.

Resumo

A Teoria de Valores Extremos desempenha um papel fundamental na modelação de acontecimentos extremos num contexto espacial. Um dos aspetos cruciais no estudo de extremos espaciais é a caracterização da dependência entre extremos em diferentes localizações, dado que estes podem não se manifestar de forma isolada.

Compreender e quantificar a dependência espacial é essencial para a análise e previsão rigorosa dos acontecimentos, permitindo uma avaliação fiável dos riscos a eles associados e o desenvolvimento de estratégias eficazes para a sua mitigação.

Neste contexto, os madogramas — coeficientes de dependência extremal enraizados na geoestatística, em particular no conceito de variograma — constituem ferramentas robustas para a quantificação da dependência nos campos max-estáveis, uma classe de modelos da Teoria de Valores Extremos particularmente adequada para o estudo de extremos espaciais.

O objetivo desta dissertação é precisamente o estudo de madogramas em campos aleatórios max-estáveis, promovendo uma abordagem interdisciplinar entre a Geoestatística e a Teoria de Valores Extremos que potencie uma modelação sólida e coesa da ocorrência de acontecimentos extremos.

Palavras-chave

Acontecimento extremo; max-estabilidade; dependência espacial; estacionariedade; madograma.

Abstract

Extreme Value Theory plays a pivotal role in modeling extreme events within a spatial framework. One of the key aspects of studying spatial extremes is characterizing the dependence between extremes occurring at different locations, given that such events may not manifest independently.

Understanding and quantifying spatial dependence is vital for the accurate analysis and forecast of such events, enabling reliable risk assessment and the development of effective mitigation strategies.

In this context, madograms — coefficients of extremal dependence rooted in geostatistics, particularly in the concept of the variogram — constitute robust tools for quantifying dependence in max-stable fields, a class of models within Extreme Value Theory that is particularly well-suited for the study of spatial extremes.

The aim of this dissertation is precisely the study of madograms in max-stable random fields, promoting an interdisciplinary approach between Geostatistics and Extreme Value Theory that fosters a solid and coherent modeling of the occurrence of extreme events.

Keywords

Extreme event; max-stability; spatial dependence; stationarity; madogram.

Conteúdo

Introdução	1
1 Fundamentos de Teoria de Valores Extremos	3
1.1 Caso univariado	4
1.2 Caso multivariado	9
1.2.1 Distribuição multivariada de valores extremos, max-estabilidade e medida expoente	10
1.2.2 Dependência entre extremos univariados	19
1.2.3 Dependência entre extremos multivariados	26
Função coeficiente extremal regional	27
2 Fundamentos de Geoestatística	33
2.1 Fundamentos dos campos aleatórios	34
2.1.1 Definições e notação	34
2.1.2 Algumas propriedades importantes	35
2.2 Funções de covariância e variogramas	39
3 Campos aleatórios max-estáveis e dependência	47
3.1 Campos aleatórios max-estáveis	47
3.2 Dependência	52
3.2.1 Função coeficiente extremal	52
3.2.2 Madogramas	54
Madograma de Matheron	55
Madograma normalizado	60
λ -madograma	64
λ -madograma multivariado	69
Madograma generalizado	75
Madograma generalizado multiregional	85
Conclusões	101
Apêndice	107

Lista de Figuras

1.1	Cauda direita de uma função de distribuição, colorida a cinza.	8
1.2	Funções densidade e de sobrevivência GEV para $\xi = 0$ (a verde), $\xi = 1$ (a vermelho) e $\xi = -1$ (a azul), com $\lambda = 0$ e $\delta = 1$	9
1.3	Representação gráfica da função de distribuição BEV G do Exemplo 2.	12
1.4	Representação gráfica da distribuição logística bivariada com $\alpha = 0.5$	18
1.5	Gráficos de dispersão de dados do modelo logístico bivariado para $\alpha = 0.05$ (à esquerda), $\alpha = 0.5$ (no meio) e $\alpha = 1$ (à direita).	19
1.6	Gráfico do índice de dependência extremal da distribuição BEV logística.	23
2.1	Semivariograma do modelo esférico com $\tau^2 = 0.2$ e $\sigma^2 = \phi = 1$	41
2.2	Gráficos das funções de correlação Whittle-Matérn (esquerda) e Exponencial Potenciada (direita), com $\phi = 1$, considerando $\nu \in \{0.5, 1.5, 2, 4\}$ para a Whittle-Matérn e $\nu \in \{0.5, 1, 1.5, 2\}$ para a Exponencial Potenciada.	44
2.3	Gráficos dos semivariogramas das famílias Whittle-Matérn (esquerda) e Exponencial Potenciada (direita), com $\phi = 1 = \sigma^2$, $\tau^2 = 0.2$, considerando $\nu \in \{0.5, 1.5, 2, 4\}$ para a Whittle-Matérn e $\nu \in \{0.5, 1, 1.5, 2\}$ para a Exponencial Potenciada.	44
2.4	Gráficos das funções de covariância das famílias Whittle-Matérn (esquerda) e Exponencial Potenciada (direita), com $\phi = 1 = \sigma^2$, $\tau^2 = 0.2$, considerando $\nu \in \{0.5, 1.5, 2, 4\}$ para a Whittle-Matérn e $\nu \in \{0.5, 1, 1.5, 2\}$ para a Exponencial Potenciada.	45
3.1	Simulação de modelo de Schlather com espaço das localizações $\mathcal{D} = [0, 10]^2$ e função de correlação exponencial para $\phi = 1$. Os valores do campo estão apresentados em escala logarítmica natural, de forma a realçar as suas variações.	50
3.2	Simulação de modelo de Smith com espaço das localizações $\mathcal{D} = [0, 10]^2$ e matriz de covariâncias $\Sigma = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}$. Os valores do campo estão apresentados em escala logarítmica natural, de forma a realçar as suas variações.	52
3.3	Função coeficiente extremal para modelo de Schlather isotrópico e estacionário com função de correlação exponencial com $\phi = 1$	54
3.4	Função coeficiente extremal de um campo de Smith isotrópico com matriz de covariâncias $\Sigma = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}$	55

3.5	Mapa da Suíça com as localizações l_i , com índices $i \in \{1, \dots, 79\}$ assinalados no mapa, das estações de medição da precipitação do conjunto de dados <i>rainfall</i> .	59
3.6	Estimativas obtidas para o madograma (gráfico à esquerda) e para a função coeficiente extremal (gráfico à direita) do conjunto de dados <i>rainfall</i> .	60
3.7	Estimativas obtidas para o F -madograma (gráfico à esquerda) e para a função coeficiente extremal (gráfico à direita) do conjunto de dados <i>rainfall</i> .	63
3.8	Estimativas do λ -madograma para o conjunto de dados <i>rainfall</i> .	68
3.9	Simulação do campo $M4$ apresentado no Exemplo 14.	72
3.10	Estimativas de $\nu_F(\lambda, \lambda, \lambda, \lambda)$, com $\lambda \in (0, 100)$, para o Conjunto A (a preto) e o Conjunto B (a azul), considerados no Exemplo 15, e limites teóricos inferior (a vermelho, igual a zero) e superior (a verde, igual a $1/(1 + \lambda/4) - 1/(1 + \lambda)$).	75
3.11	Coordenadas dos pontos $x_i, i \in \{1, \dots, 16\}$, coloridos a laranja os indexados em $I_1 = \{1, \dots, 8\}$ e a azul os em $I_2 = \{9, \dots, 16\}$.	78
3.12	Estimativas de $\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda, \lambda))$, com $\lambda \in (0, 100)$, para o Exemplo 17.	82
3.13	Estimativas de $\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda/3, (1 - \lambda)/2))$, com $\lambda \in (0, 1)$, para o Exemplo 17.	83
3.14	Estimativas de $\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2))$, com $\lambda_1 \in (0, 100)$ e $\lambda_2 \in \{0.5, 1, 8, 50, 100\}$, para o Exemplo 17.	84
3.15	Estimativas de $\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2))$, com $\lambda_1, \lambda_2 \in (0, 100)$, para o Exemplo 17.	84
3.16	Coordenadas dos pontos indexados na partição \mathcal{I} do Exemplo 18.	90
3.17	Estimativas de $\hat{\nu}_F(\mathbf{Z}, \{I_1, I_2, I_3\}, (\lambda, \lambda, \lambda))$ para configuração A do Exemplo 19 com $\lambda \in (0, 100)$.	98
3.18	Estimativas de $\hat{\nu}_F(\mathbf{Z}, \{I_1, I_2, I_3\}, (\lambda, \lambda, \lambda))$ para configuração B do Exemplo 19 com $\lambda \in (0, 100)$.	99
3.19	Estimativas de $\hat{\nu}_F(\mathbf{Z}, \{I_1, I_2, I_3\}, (\lambda, \lambda, \lambda))$ para configuração C do Exemplo 19 com $\lambda \in (0, 100)$.	99

Lista de Tabelas

2.1	Famílias paramétricas de semivariogramas isotrópicos válidos obtidos a partir das funções de correlação definidas anteriormente.	45
2.2	Famílias paramétricas de funções de covariância isotrópicas válidas obtidas a partir dos semivariogramas da Tabela 2.1.	46
3.1	Longitudes, latitudes e altitudes do subconjunto em estudo no Exemplo 15. . .	74
3.2	Estimativas do madograma e da função coeficiente extremal para os conjuntos A e B	74
3.3	Longitudes, latitudes e altitudes das localizações selecionadas no Exemplo 17.	82
3.4	Longitudes, latitudes e altitudes das estações 34, 39, 61, 70 e 77.	97

Introdução

Ao longo dos últimos anos, tem-se verificado um aumento significativo na frequência e intensidade de muitos acontecimentos extremos climatológicos, como precipitações e temperaturas elevadas, num número crescente de localizações do mundo. Estudos científicos alertam que esta tendência perpetuar-se-á nos próximos anos, o que torna os acontecimentos extremos climatológicos uma ameaça crescente para o bem-estar das populações e ecossistemas de todo o planeta, dados os seus fortes impactos socioeconómicos e ecológicos.

O estudo de acontecimentos extremos num contexto espacial tornou-se, portanto, imperativo. Apesar de as ferramentas para a modelação de extremos univariados e multivariados estarem bem consolidadas, a sua extensão para modelar extremos espaciais constitui ainda uma área de investigação em desenvolvimento.

Nesta, os campos aleatórios max-estáveis emergem como modelos centrais para o estudo de valores extremos no espaço, destacando-se em aplicações a casos práticos (Smith (1990) [44], Coles e Tawn (1996) [4], Schlather e Tawn (2003) [40], Buishand et al. (2008) [2], entre outros). A existência de uma infinidade de campos max-estáveis, ilustrada, por exemplo, pela representação espectral proposta por de Haan (1984) [8], releva o desafio prático de escolher modelos que capturem de forma adequada a dependência extremal.

A necessidade de modelar fenómenos naturais, nomeadamente os meteorológicos, conduziu ao desenvolvimento de métodos para definir e simular classes particulares de campos max-estáveis (Brown e Resnick (1977) [1], Smith (1990) [44], Schlather (2002) [39], Kabluchko (2009) [26]). Contudo, 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. Posto isto, a quantificação da dependência em campos max-estáveis é um tópico de grande relevância, tendo vindo a despertar o interesse de muitos investigadores.

Com este trabalho pretende-se apresentar alguns coeficientes de dependência para campos max-estáveis que tiveram origem no conceito de variograma, utilizado em Geoestatística, estudar as suas propriedades, relações com outras medidas de dependência e, por último, a sua aplicação a dados reais.

Começamos, então, por, no Capítulo 1, introduzir a abordagem clássica para o estudo de acontecimentos extremos: a análise do comportamento limite do máximo, linearmente normalizado, de uma sucessão de variáveis aleatórias reais independentes e identicamente distribuídas. Este enquadramento, correspondente ao caso univariado, permite a modelação de um único acontecimento extremo. Posteriormente, alarga-se a análise ao caso multivariado, considerando sucessões de vetores aleatórios, o que possibilita o estudo conjunto de múlti-

plos acontecimentos extremos. Em ambos os contextos, verifica-se que o comportamento assintótico dos máximos é bem modelado por distribuições max-estáveis. No cenário multivariado, explora-se ainda a dependência, com a introdução de medidas que estabelecerão um elo conceptual com os coeficientes de dependência a desenvolver no contexto dos campos aleatórios max-estáveis.

No Capítulo 2, apresentam-se os objetos de estudo da Geoestatística — os campos aleatórios —, bem como o conceito de estacionariedade, essencial para a sua modelação. Introduzem-se as definições de função covariância e semivariograma, fundamentais para a análise da dependência espacial em campos aleatórios, explorando-se as relações entre ambas e as respetivas propriedades.

No Capítulo 3, introduzem-se os campos aleatórios max-estáveis, que constituem uma extensão natural, no domínio espacial, dos vetores com distribuições max-estáveis tratados no Capítulo 1. Ressalta-se que, para estes, as ferramentas geoestatísticas convencionais se revelam inadequadas para a avaliação da dependência espacial. Contudo, com inspiração na definição do semivariograma, obtêm-se coeficientes alternativos — os madogramas — que permitem quantificar a dependência entre acontecimentos extremos em diferentes localizações. São apresentadas as definições de alguns, bem como as suas propriedades e ligações a medidas de dependência introduzidas no Capítulo 1, sendo ilustrados com a aplicação a um conjunto de dados reais, com o auxílio do *software R*.

Concluimos com uma reflexão crítica sobre o trabalho desenvolvido e com a identificação de possíveis direções para investigação futura.

Capítulo 1

Fundamentos de Teoria de Valores Extremos

Em numerosos domínios científicos, privilegia-se a análise do comportamento central das distribuições de dados. Contudo, esta abordagem revela-se insuficiente — e até perigosa — em determinados contextos. As grandes estruturas de engenharia não ruem sob condições médias; os mercados financeiros não colapsam em dias banais; os ecossistemas não entram em falência perante variações triviais. É precisamente nesses contextos — onde os extremos determinam catástrofes, colapsos ou recordes históricos — que a Teoria de Valores Extremos se impõe como ferramenta essencial. Por essa razão, a TVE tem vindo a assumir uma crescente relevância no seio da comunidade científica, como o demonstra a sua ampla aplicação em áreas tão diversas quanto a Hidrologia, a Climatologia, a Fiabilidade, a Resistência de Materiais, a Engenharia Estrutural, o Controlo de Qualidade, as Finanças e os Seguros. Enraizada na década de 1920, a TVE, enquanto ramo de Probabilidades e Estatística, tem como cerne a compreensão do comportamento dos acontecimentos extremos — acontecimentos cuja probabilidade de ocorrência é muito pequena e nos quais as variáveis em causa atingem valores excepcionalmente elevados ou reduzidos. No que se segue, restringir-nos-emos ao estudo do comportamento do máximo de uma sucessão de variáveis aleatórias independentes e identicamente distribuídas. Note-se que este é facilmente transponível para o do mínimo, atendendo à relação $\bigvee_{i=1}^n Z_i = -\bigwedge_{i=1}^n -Z_i$, para as variáveis aleatórias Z_1, \dots, Z_n , $n \in \mathbb{N}$, sendo que ‘ \bigvee ’ e ‘ \bigwedge ’ denotam, respetivamente, o máximo e o mínimo das variáveis Z_1, \dots, Z_n . Começamos por abordar o caso univariado, na Secção 1.1. Contudo, muitos fenómenos de interesse, nomeadamente em contextos ambientais, apresentam um carácter inerentemente multivariado. Por exemplo, na análise da qualidade do ar em determinada região, o nível de poluição do ar depende dos níveis de vários poluentes e da dependência entre eles, o que exige uma abordagem conjunta dos seus extremos. Esta motivação torna inevitável a transição para a TVE multivariada, tema da Secção 1.2.

1.1 Caso univariado

Dada uma sucessão de variáveis aleatórias reais $\{Z_n\}_{n \in \mathbb{N}}$ independentes e identicamente distribuídas (i.i.d.) com função de distribuição (f.d.) F , denote-se por M_n o máximo das n primeiras, com $n \in \mathbb{N}$, *i.e.*,

$$M_n := \bigvee_{i=1}^n Z_i.$$

A distribuição exata de M_n é dada por

$$\begin{aligned} P(M_n \leq x) &= P(Z_1 \leq z, \dots, Z_n \leq z) \\ &= P^n(Z_1 \leq z) \\ &= F^n(z), \end{aligned} \tag{1.1}$$

pelo que, na prática, (1.1) só tem interesse se a f.d. F for conhecida.

Em geral, a distribuição F é desconhecida. Mesmo que se estime a partir de uma realização da amostra aleatória (a.a.) (Z_1, \dots, Z_n) , o erro associado pode propagar-se de forma significativa ao considerar a distribuição exata do máximo, F^n , sobretudo quando o tamanho da amostra, n , for suficientemente grande. Assim, revela-se mais viável estudar o comportamento assintótico da f.d. do máximo do que da sua forma exata. Então, fazendo n tender para infinito em (1.1), obtém-se a f.d. assintótica de M_n

$$\lim_{n \rightarrow +\infty} F^n(z) = \begin{cases} 0, & \text{se } z < z^F, \\ 1, & \text{se } z \geq z^F, \end{cases} \tag{1.2}$$

onde z^F denota o limite superior do suporte de F , *i.e.*,

$$z^F = \sup \{z : F(z) < 1\}.$$

Donde se conclui que a f.d. assintótica de M_n é degenerada e, portanto, desprovida de interesse. Posto isto, e tendo por inspiração a Teoria do Limite Central, procede-se à normalização linear do máximo M_n de forma a perceber se, assim, se obtém uma f.d. limite não degenerada, *i.e.*, de forma a perceber se a condição

$$P\left(\frac{M_n - \lambda_n}{\delta_n} \leq z\right) = F^n(\delta_n z + \lambda_n) \xrightarrow{n \rightarrow +\infty} G(z), \quad \forall z \in \mathcal{C}(G), \tag{1.3}$$

é válida para algumas constantes reais $\delta_n > 0$ e λ_n , $n \in \mathbb{N}$, e para alguma f.d. G não degenerada, onde $\mathcal{C}(G)$ denota o conjunto dos pontos de continuidade de G .

Quando a condição (1.3) é válida, as constantes δ_n e λ_n designam-se **coeficientes de atração** de F para G e diz-se que F pertence ao **max-domínio de atração** de G , escrevendo-se, simbolicamente, $F \in \mathcal{D}_{\mathcal{M}}(G)$, sendo que $\mathcal{D}_{\mathcal{M}}(G)$ denota, assim, o max-domínio de atração de G definido por

$$\mathcal{D}_{\mathcal{M}}(G) := \left\{ F : \exists \delta_n \in \mathbb{R}^+, \lambda_n \in \mathbb{R} \mid F^n(\delta_n z + \lambda_n) \xrightarrow{n \rightarrow +\infty} G(z), \forall z \in \mathcal{C}(G) \right\}.$$

Nesse caso, diz-se que a sucessão de funções de distribuição de máximos normalizados converge fracamente para a f.d. G , ou equivalentemente, que a sucessão de máximos normalizados converge em distribuição para uma variável com f.d. G .

A convergência em (1.3) levanta duas questões principais: a primeira é determinar as formas possíveis das f.d. não degeneradas G que satisfaçam a relação (1.3), considerando sucessões apropriadas de constantes; a segunda é caracterizar o domínio de atração de G , ou seja, identificar a classe de f.d. F para as quais (1.3) é válida.

Antes de apresentarmos a resposta à primeira questão, introduzimos o conceito de funções de distribuição do mesmo tipo.

Definição 1. *Duas funções de distribuição H_1 e H_2 são **do mesmo tipo** se existirem constantes reais $\delta > 0$ e λ tais que*

$$H_1(\delta z + \lambda) = H_2(z), \quad (1.4)$$

para todo o z real.

Note-se que ter (1.4) é equivalente a ter

$$H_1(z) = H_2\left(\frac{z - \lambda}{\delta}\right), \quad (1.5)$$

ou seja, duas f.d. H_1 e H_2 são do mesmo tipo se são a mesma, a menos de um parâmetro de localização $\lambda \in \mathbb{R}$ e de um parâmetro de escala $\delta \in \mathbb{R}^+$.

Tendo, agora, a definição anterior presente, enunciamos o resultado que espoletou grande parte da investigação neste tema, designado *Teorema dos Tipos Extremos* e formalmente demonstrado por Gnedenko (1943) [22]. Este estabelece a classe de distribuições a que G em (1.3) pertence ou, dito de outra forma, estabelece o domínio de atração a que pertence a f.d. F comum às variáveis da sucessão.

Teorema 1 (Teorema dos Tipos Extremos). *Seja $\{Z_n\}_{n \in \mathbb{N}}$ uma sucessão de variáveis aleatórias reais independentes e identicamente distribuídas com função de distribuição F , verificando (1.3) para algumas sucessões de constantes reais $\{\delta_n > 0\}_{n \in \mathbb{N}}$ e $\{\lambda_n\}_{n \in \mathbb{N}}$ e para alguma f.d. não degenerada G . Então, G é do tipo de uma das três distribuições seguintes*

$$\Lambda(z) = \exp(-\exp(-z)), \quad z \in \mathbb{R}; \quad (1.6)$$

$$\Phi_\alpha(z) = \begin{cases} 0, & \text{se } z \leq 0, \\ \exp(-z^{-\alpha}), & \text{se } z > 0, \quad \alpha > 0; \end{cases} \quad (1.7)$$

$$\Psi_\alpha(z) = \begin{cases} \exp(-(-z)^\alpha), & \text{se } z < 0, \quad \alpha > 0, \\ 1, & \text{se } z \geq 0. \end{cases} \quad (1.8)$$

Este Teorema dá-nos, então, a resposta à primeira questão levantada, afirmando que se a condição (1.3) é verificada para algumas sucessões de constantes reais $\{\delta_n > 0\}_{n \in \mathbb{N}}$ e $\{\lambda_n\}_{n \in \mathbb{N}}$ e para alguma f.d. não degenerada G , então G é do tipo de uma das três f.d. especificadas anteriormente.

Se for do tipo de (1.6), tem expressão

$$\Lambda_{\lambda, \delta}(z) = \exp\left(-\exp\left(-\frac{z-\lambda}{\delta}\right)\right), \quad z \in \mathbb{R}; \quad (1.9)$$

e diz-se que é uma **distribuição Gumbel**, ou uma distribuição do Tipo I; se for do tipo de (1.7), então diz-se uma **distribuição Fréchet**, ou do Tipo II, e tem expressão

$$\Phi_{\lambda, \delta, \alpha}(z) = \begin{cases} 0, & \text{se } z \leq \lambda, \\ \exp\left(-\left(\frac{z-\lambda}{\delta}\right)^{-\alpha}\right), & \text{se } z > \lambda, \quad \alpha > 0; \end{cases} \quad (1.10)$$

enquanto, se for do tipo de (1.8), diz-se que é uma **distribuição Weibull**, ou do Tipo III, tendo como expressão

$$\Psi_{\lambda, \delta, \alpha}(z) = \begin{cases} \exp\left(-\left(-\frac{z-\lambda}{\delta}\right)^\alpha\right), & \text{se } z < \lambda, \quad \alpha > 0, \\ 1, & \text{se } z \geq \lambda; \end{cases} \quad (1.11)$$

com $\lambda \in \mathbb{R}$ e $\delta \in \mathbb{R}^+$.

Estas distribuições — (1.9), (1.10) e (1.11) — são conhecidas como **distribuições de valores extremos** e coincidem precisamente com as distribuições max-estáveis não degeneradas.

Proposição 1. *Uma distribuição é de valores extremos se e só se é max-estável não degenerada.*

Uma distribuição H é max-estável quando é a mesma, a menos de parâmetros de localização e escala, que a distribuição H^n do máximo da amostra a ela subjacente, ou seja, H é do mesmo tipo que H^n , para todo o inteiro positivo n .

Definição 2. Uma função de distribuição H diz-se **max-estável** se existirem constantes reais $\tilde{\delta}_n > 0$ e $\tilde{\lambda}_n$ tais que

$$H^n(\tilde{\delta}_n z + \tilde{\lambda}_n) = H(z), \quad (1.12)$$

para todo o inteiro positivo n e para todo o $z \in \mathbb{R}$.

As distribuições max-estáveis não degeneradas, ou equivalentemente, as distribuições de valores extremos, podem ser representadas numa forma paramétrica única, conhecida como **forma de von-Mises Jenkinson** ou **distribuição Generalizada de Valores Extremos** (GEV, do inglês ‘Generalized Extreme Value’), de expressão

$$G_{\lambda,\delta,\xi}(z) = \begin{cases} \exp\left(-\left(1 + \xi \frac{z-\lambda}{\delta}\right)^{-1/\xi}\right), & \text{se } 1 + \xi \frac{z-\lambda}{\delta} \in \mathbb{R}^+ \text{ e } \xi \neq 0, \\ \exp\left(-\exp\left(-\frac{z-\lambda}{\delta}\right)\right), & \text{se } z \in \mathbb{R} \text{ e } \xi = 0; \end{cases} \quad (1.13)$$

sendo $\lambda \in \mathbb{R}$, $\delta \in \mathbb{R}^+$ e $\xi \in \mathbb{R}$ parâmetros de localização, escala e forma, respetivamente. O parâmetro ξ é designado índice de valores extremos e o seu valor determina qual o Tipo da distribuição limite G em (1.3), facilitando a implementação da teoria em aplicações estatísticas.

— Se $\xi = 0$ e considerando os limites de (1.13) quando ξ tende para 0 por valores superiores e inferiores, obtém-se que G é uma distribuição do Tipo I, ou Gumbel, mais precisamente

$$G_{\lambda,\delta,0}(z) = \lim_{\xi \rightarrow 0^+} G_{\lambda,\delta,\xi}(z) = \lim_{\xi \rightarrow 0^-} G_{\lambda,\delta,\xi}(z) = \Lambda_{\lambda,\delta}(z), \quad z \in \mathbb{R}.$$

— Se $\xi > 0$, tem-se que G é uma distribuição do Tipo II, ou Fréchet, com

$$G_{\lambda,\delta,\xi}(z) = \Phi_{\lambda-\delta/\xi,\delta/\xi,1/\xi}(z), \quad 1 + \xi \frac{z-\lambda}{\delta} > 0.$$

— Se $\xi < 0$, então G é uma distribuição do Tipo III, ou Weibull, com

$$G_{\lambda,\delta,\xi}(z) = \Psi_{\lambda-\delta/\xi,-\delta/\xi,-1/\xi}(z), \quad 1 + \xi \frac{z-\lambda}{\delta} > 0.$$

Além disso, ξ também pode ser designado índice de cauda, pois está relacionado com o peso da cauda superior da f.d. F atraída para G , no sentido de (1.3).

Definição 3. A **cauda superior** da função de distribuição F , da variável Z , também designada função de sobrevivência de Z , é dada por

$$\bar{F}(z) = 1 - F(z) = P(Z > z),$$

para $z \in \mathbb{R}$.

A ilustração gráfica deste conceito é apresentada na Figura 1.1.

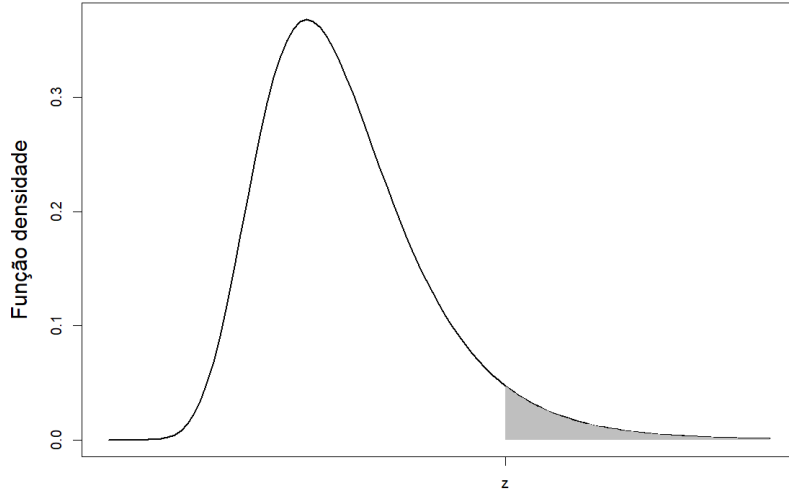


Figura 1.1: Cauda direita de uma função de distribuição, colorida a cinza.

O valor de ξ espelha o comportamento da cauda superior de F , quando z tende para o limite superior do suporte, z^F .

— Se $\xi = 0$, a cauda superior de F é do tipo exponencial e o limite superior de F pode ser finito ou infinito. A título de exemplo têm-se as distribuições Gumbel, Exponencial e de Gauss.

— Se $\xi > 0$, a cauda superior de F é pesada, do tipo polinomial e o limite superior do suporte de F é infinito. As distribuições Fréchet, T-Student e Pareto são exemplos deste caso.

— Se $\xi < 0$, a cauda é curta ou leve e o limite superior do suporte é finito. A Weibull, Uniforme e Beta são exemplos de tais distribuições.

Estes três casos são ilustrados na Figura 1.2¹.

Relativamente à questão referente à caracterização dos domínios de atração de cada um dos três tipos de distribuições de valores extremos, prova-se que cada domínio de atração é não vazio, pois qualquer uma das distribuições de valores extremos pertence ao seu próprio domínio de atração, e que os domínios de atração são disjuntos (Leadbetter et al.(1983) [27]). Além disso, as condições para que uma f.d. F pertença a algum dos três domínios de atração incidem especialmente sobre o comportamento assintótico da cauda de F (cf, por exemplo, Gnedenko (1943) [22] e de Haan (1970) [7]).

Finalizamos esta secção mostrando que, efetivamente, a distribuição Fréchet unitária pertence ao domínio de atração de ela própria.

¹Note-se que a função densidade de probabilidade da GEV apresentada na Figura 1.2 para diferentes valores de ξ , obtém-se derivando (1.13) em ordem a z , sendo dada por

$$g_{\lambda,\delta,\xi}(z) = \begin{cases} \frac{1}{\delta} \left(1 + \xi \frac{z-\lambda}{\delta}\right)^{-1/\xi-1} \exp\left(-\left(1 + \xi \frac{z-\lambda}{\delta}\right)^{-1/\xi}\right), & \text{se } 1 + \xi \frac{z-\lambda}{\delta} \in \mathbb{R}^+ \text{ e } \xi \neq 0, \\ \frac{1}{\delta} \exp\left(-\frac{z-\lambda}{\delta}\right) \exp\left(-\exp\left(-\frac{z-\lambda}{\delta}\right)\right), & \text{se } z \in \mathbb{R} \text{ e } \xi = 0; \end{cases}$$

com $\lambda \in \mathbb{R}$ e $\delta \in \mathbb{R}^+$.

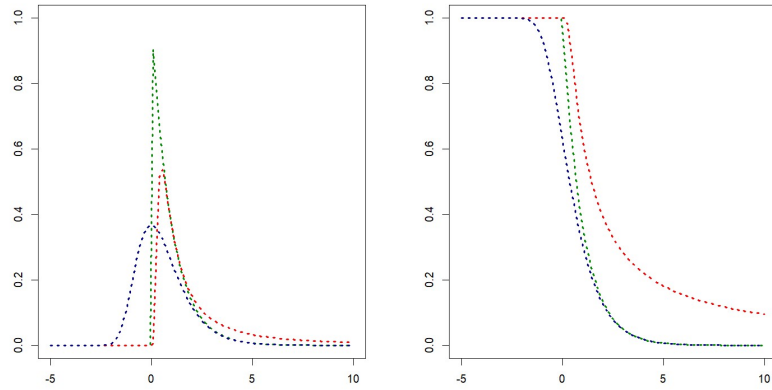


Figura 1.2: Funções densidade e de sobrevivência GEV para $\xi = 0$ (a verde), $\xi = 1$ (a vermelho) e $\xi = -1$ (a azul), com $\lambda = 0$ e $\delta = 1$.

Exemplo 1. Considere-se uma sucessão de v.a. $\{Z_n\}_{n \in \mathbb{N}}$ i.i.d. com f.d. Fréchet unitária, i.e., com f.d. $F(z) = \exp(-1/z) = \Phi_{0,1,1}(z)$, para $z > 0$. Então, para $\delta_n = n$ e $\lambda_n = 0$, a condição (1.3) verifica-se e obtém-se como distribuição limite G a distribuição Fréchet unitária:

$$\begin{aligned}
 P\left(\frac{M_n - \lambda_n}{\delta_n} \leq z\right) &= P\left(\frac{M_n}{n} \leq z\right) \\
 &= F^n(nz) \\
 &= \left(\exp\left(-\frac{1}{nz}\right)\right)^n \\
 &= \exp\left(-\frac{1}{z}\right), \quad z > 0, \quad \forall n \in \mathbb{N}.
 \end{aligned}$$

1.2 Caso multivariado

A Teoria de Valores Extremos para sucessões multidimensionais apresenta vários desafios aos investigadores, devido às dificuldades específicas de modelação e extensão dos resultados já disponíveis para sucessões de variáveis aleatórias reais. Surgem novas questões características do contexto multivariado, como, por exemplo, a definição de máximo, dada a inexistência de uma maneira óbvia de ordenar observações. A abordagem clássica tem sido considerar os máximos de uma amostra multivariada como um vetor cujas componentes são os máximos marginais, como apresentaremos na próxima subsecção. Estabelecida a definição de máximo multivariado avançamos — similarmente ao caso univariado — para a abordagem clássica, que consiste no estudo do comportamento assintótico da distribuição do máximo linearmente normalizado no contexto de vetores i.i.d.. Verifica-se que este é bem modelado por uma distribuição multivariada de valores extremos (MEV), sendo que a classe de distribuições MEV coincide com a classe de distribuições max-estáveis, à semelhança do caso univariado. Contudo, contrariamente ao contexto univariado, esta classe não admite uma

representação paramétrica finita, o que constitui outra dificuldade inerente a esta extensão da teoria de uma para várias variáveis.

Por último, na modelação de extremos multivariados, é crucial estudar a dependência entre as observações extremas. Uma abordagem a este problema — e com a qual finalizamos esta secção — consiste em medir a dependência na distribuição limite do vetor de máximos linearmente normalizado utilizando diferentes medidas de dependência existentes na literatura.

1.2.1 Distribuição multivariada de valores extremos, max-estabilidade e medida expoente

Seja $\{\mathbf{Z}_n = (Z_{n,1}, \dots, Z_{n,k})\}_{n \in \mathbb{N}}$ uma sucessão de vetores aleatórios reais (ve.a.r.) definidos sobre um mesmo espaço de probabilidade (Ω, \mathcal{A}, P) , independentes e com função de distribuição comum $F(\mathbf{z}) = F(z_1, \dots, z_k)$, para $\mathbf{z} \in \mathbb{R}^k$ e k um inteiro maior do que um.

Tal como referimos anteriormente, a primeira dificuldade encontrada na passagem do contexto unidimensional para o contexto multidimensional foi a definição de valor extremo multivariado. Esta definição depende da relação de ordem considerada no espaço real k -dimensional. A abordagem clássica consiste em considerar o máximo multivariado como o vetor de componentes máximas, *i.e.*,

$$\mathbf{M}_n := (M_{n,1}, \dots, M_{n,k}), \text{ onde } M_{n,j} := \bigvee_{i=1}^n Z_{i,j}, \text{ com } j = 1, \dots, k,$$

e a seguinte relação de ordem

$$\mathbf{x} \leq \mathbf{y} \text{ se e só se } x_j \leq y_j, j = 1, \dots, k.$$

As operações entre vetores em \mathbb{R}^k também serão consideradas componente a componente.²

À semelhança do que sucede na teoria univariada, a análise da distribuição de \mathbf{M}_n linearmente normalizado, para n suficientemente grande, é o ponto de partida para a teoria de valores extremos multivariados:

Seja $\{\mathbf{Z}_n = (Z_{n,1}, \dots, Z_{n,k})\}_{n \in \mathbb{N}}$ uma sucessão de ve.a.r. i.i.d. com f.d. F . Se existirem sucessões reais $\{\boldsymbol{\delta}_n = (\delta_{n,1}, \dots, \delta_{n,k})\}_{n \in \mathbb{N}}$, com $\boldsymbol{\delta}_n > \mathbf{0}$, e $\{\boldsymbol{\lambda}_n = (\lambda_{n,1}, \dots, \lambda_{n,k})\}_{n \in \mathbb{N}}$ tais que

$$P\left(\frac{\mathbf{M}_n - \boldsymbol{\lambda}_n}{\boldsymbol{\delta}_n} \leq \mathbf{z}\right) = F^n(\boldsymbol{\delta}_n \mathbf{z} + \boldsymbol{\lambda}_n) \xrightarrow[n \rightarrow +\infty]{} G(\mathbf{z}), \quad \forall \mathbf{z} \in \mathcal{C}(G), \quad (1.14)$$

para alguma f.d. G com marginais univariadas não degeneradas, então diz-se que G é uma f.d. ***k*-variada de valores extremos** e que F pertence ao seu *max-domínio de atracção*

²Dados $\mathbf{x} = (x_1, \dots, x_k)$, $\mathbf{y} = (y_1, \dots, y_k)$, $\mathbf{z} = (z_1, \dots, z_k) \in \mathbb{R}^k$, com $\mathbf{x} \neq \mathbf{0} = (0, \dots, 0)$, e $a \in \mathbb{R}$, tem-se

$$\mathbf{x}^{-1} \mathbf{y} + a \mathbf{z} = (x_1^{-1} y_1 + a z_1, \dots, x_k^{-1} y_k + a z_k).$$

(k -variado). Simbolicamente, escrevemos $F \in \mathcal{D}_{\mathcal{M}}(G)$, sendo que

$$\mathcal{D}_{\mathcal{M}}(G) = \{F : \exists \{\delta_n > \mathbf{0}\}_{n \in \mathbb{N}}, \{\lambda_n\}_{n \in \mathbb{N}} \mid F^n(\delta_n z + \lambda_n) \xrightarrow{n \rightarrow +\infty} G(z), \forall z \in \mathcal{C}(G)\},$$

onde $\mathcal{C}(G)$ denota o conjunto dos pontos de continuidade de G .

A classe de f.d. k -variadas G que surgem como limite em (1.14) coincide com a classe de f.d. k -variadas max-estáveis com f.d. marginais univariadas não degeneradas, sendo a definição de f.d. k -variada max-estável apresentada de seguida.

Definição 4. *Uma função de distribuição k -variada G diz-se **max-estável** se, para qualquer inteiro positivo n , existem constantes $\tilde{\delta}_n = (\tilde{\delta}_{n,1}, \dots, \tilde{\delta}_{n,k})$, com $\tilde{\delta}_n > \mathbf{0}$, e $\tilde{\lambda}_n = (\tilde{\lambda}_{n,1}, \dots, \tilde{\lambda}_{n,k}) \in \mathbb{R}^k$, tais que a equação*

$$G^n(\tilde{\delta}_n z + \tilde{\lambda}_n) = G(z) \quad (1.15)$$

é válida para cada $z = (z_1, \dots, z_k) \in \mathbb{R}^k$.

Formalizamos a afirmação feita anteriormente no resultado que se segue.

Proposição 2. *Uma função de distribuição G multivariada é de valores extremos se e somente se é max-estável com marginais univariadas não degeneradas.*

No exemplo que se segue, recorrendo a (1.14), apresentamos uma distribuição bivariada de valores extremos (BEV).

Exemplo 2. *Seja $\{(Z_{n,1}, Z_{n,2})\}_{n \in \mathbb{N}}$ uma sucessão de ve.a. com função de Mardia*

$$F(z_1, z_2) = 1 - \exp(-z_1) - \exp(-z_2) + [\exp(z_1) + \exp(z_2) - 1]^{-1}, \quad z_1, z_2 \in \mathbb{R}^+.$$

F tem distribuições marginais exponenciais e, considerando as sucessões $\{\delta_n = (1, 1)\}_{n \in \mathbb{N}}$ e $\{\lambda_n = (\ln n, \ln n)\}_{n \in \mathbb{N}}$, tem-se

$$F^n(z_1 + \ln n, z_2 + \ln n) = \left\{ 1 + \frac{-\exp(-z_1) - \exp(-z_2) + [\exp(z_1) + \exp(z_2) - 1/n]^{-1}}{n} \right\}^n$$

$$\xrightarrow{n \rightarrow +\infty} G(z_1, z_2) = \exp \left\{ -\exp(-z_1) - \exp(-z_2) + [\exp(z_1) + \exp(z_2)]^{-1} \right\},$$

onde a função de distribuição G tem distribuições marginais Gumbel:

$$\lim_{z_2 \rightarrow +\infty} G(z_1, z_2) = \Lambda_{0,1}(z_1)$$

e

$$\lim_{z_1 \rightarrow +\infty} G(z_1, z_2) = \Lambda_{0,1}(z_2).$$

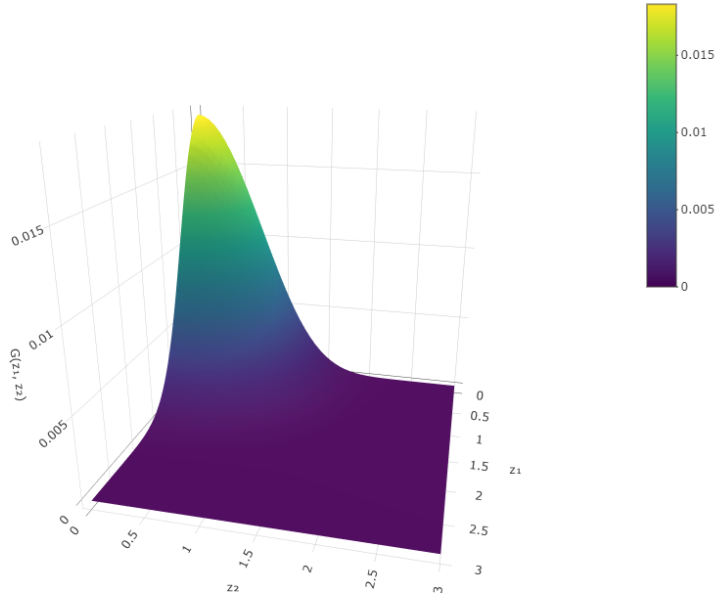


Figura 1.3: Representação gráfica da função de distribuição BEV G do Exemplo 2.

Então, como se verifica (1.14) e a função G tem margens não degeneradas, concluímos que G é uma distribuição BEV. Logo, pela proposição anterior, é max-estável. De facto, para qualquer inteiro $n \geq 1$, $\tilde{\delta}_n = (1, 1)$ e $\tilde{\lambda}_n = (\ln n, \ln n)$, tem-se

$$\begin{aligned}
 & G^n(z_1 + \ln n, z_2 + \ln n) \\
 &= \left\{ \exp[-\exp(-z_1 - \ln n)] \exp[-\exp(-z_2 - \ln n)] \right. \\
 &\quad \left. \exp[\exp(z_1 + \ln n) + \exp(z_2 + \ln n)]^{-1} \right\}^n \\
 &= \left\{ \exp[-n^{-1} \exp(-z_1)] \exp[-n^{-1} \exp(-z_2)] \exp[n^{-1} (\exp(z_1) + \exp(z_2))^{-1}] \right\}^n \\
 &= \exp[-\exp(-z_1)] \exp[-\exp(-z_2)] \exp[\exp(z_1) + \exp(z_2)]^{-1} \\
 &= G(z_1, z_2).
 \end{aligned}$$

O gráfico de G é apresentado na Figura 1.3.

Em de Haan e Resnick (1977) [10], estabelecem-se condições necessárias e suficientes para a convergência fraca da sucessão de f.d. de máximos de uma sucessão de vetores i.i.d., linearmente normalizados, para uma f.d. G com distribuições marginais univariadas não degeneradas e caracterizam-se os max-domínios de atração.

A convergência fraca de uma sucessão de funções de distribuição multivariadas para uma f.d. com distribuições marginais contínuas é equivalente à convergência das f.d. marginais univariadas e da estrutura de dependência (Deheuvels (1984) [12], Galambos (1987) [21]). Em particular, a convergência fraca da sucessão de distribuições k -variadas $\{F^n\}_{n \in \mathbb{N}}$ para G ,

em (1.14), é equivalente a:

1. convergência de cada distribuição marginal univariada F_j^n de F^n para a correspondente distribuição marginal G_j de G , $j \in \{1, \dots, k\}$;
2. convergência da estrutura de dependência de F^n para a estrutura de dependência de G ;

já que as distribuições marginais univariadas de G são contínuas, como veremos de seguida. Assim sendo, a estrutura de dependência de F^n , $n \in \mathbb{N}$, e as suas distribuições marginais podem ser aproximadas pela estrutura de dependência de G e pelas distribuições marginais desta, respetivamente.

Segundo Sklar 2 (1959) [43], uma f.d. multivariada é caracterizada pelas f.d. das margens univariadas e pela sua função cópula, que descreve a estrutura de dependência destas. Então, pelo exposto, a caracterização de $\{F^n\}_{n \in \mathbb{N}}$ pode ser feita à custa da caracterização de G , *i.e.*, através das f.d. das margens univariadas subjacentes a G , bem como da sua função cópula.

Uma função cópula $C : [0, 1]^k \rightarrow [0, 1]$ é a restrição a $[0, 1]^k$ de uma função de distribuição k -variada com distribuições marginais univariadas uniformes sobre $[0, 1]$, *i.e.*,

$$C(u_1, \dots, u_k) = P(U_1 \leq u_1, \dots, U_k \leq u_k), \quad 0 \leq u_j \leq 1,$$

com $U_j \sim U([0, 1])$, $j \in \{1, \dots, k\}$. O próximo teorema, de Sklar (1959) [43], garante a existência e unicidade da função cópula.

Teorema 2. *Seja F uma função de distribuição k -variada com funções de distribuição marginais F_1, \dots, F_k . Então existe uma cópula k -dimensional $C_F(\cdot)$ tal que*

$$F(z_1, \dots, z_k) = C_F(F_1(z_1), \dots, F_k(z_k)),$$

com $z_1, \dots, z_k \in \mathbb{R}$. Se as distribuições F_j , $j \in \{1, \dots, k\}$, são contínuas, então $C_F(\cdot)$ é única e dada por

$$C_F(u_1, \dots, u_k) = F(F_1^{-1}(u_1), \dots, F_k^{-1}(u_k)),$$

com $u_1, \dots, u_k \in [0, 1]^k$.

No próximo lema é enunciada uma das propriedades mais importantes da função cópula (Hsing (1989) [25]). Esta será muito útil para a demonstração do resultado seguinte.

Lema 1. *Sejam F e G funções de distribuição k -dimensionais com funções de distribuição marginais univariadas F_j e G_j , $j \in \{1, \dots, k\}$, respetivamente. Se $F_j(\mathbb{R}) = G_j(\mathbb{R})$, para $j \in \{1, \dots, k\}$, e se existirem funções T_j , $j \in \{1, \dots, k\}$, tais que*

$$F(z_1, \dots, z_k) = G(T_1(z_1), \dots, T_k(z_k)), \quad \forall z_1, \dots, z_k \in \mathbb{R}, \quad (1.16)$$

então $C_F \equiv C_G$.

Apesar de a função cópula ser comumente utilizada na literatura para descrever a estrutura de dependência de distribuições multivariadas de valores extremos, nesta dissertação recorreremos à medida expoente, relacionada a uma medida espectral definida sobre o simplex $(k - 1)$ -dimensional, devido à sua conexão com os madogramas introduzidos no Capítulo 3. A função cópula foi introduzida apenas como ferramenta auxiliar para a demonstração de um resultado necessário ao desenvolvimento deste trabalho, apresentado na próxima posição.

No que diz respeito às distribuições marginais univariadas de G , considerando, na equação (1.14), $z_i \rightarrow +\infty$, para $i \in \{1, \dots, k\} \setminus \{j\}$, obtém-se

$$F_j^n(\delta_{n,j} z_j + \lambda_{n,j}) \xrightarrow{n \rightarrow +\infty} G_j(z_j), \quad j \in \{1, \dots, k\}, \quad \forall z_j \in \mathcal{C}(G_j), \quad (1.17)$$

onde F_j e G_j são as j -ésimas distribuições marginais univariadas de F e G , respetivamente. Pelo Teorema dos Tipos Extremos 1, segue-se que cada distribuição marginal univariada de G , G_j , $j \in \{1, \dots, k\}$, é uma das três distribuições univariadas de valores extremos — Gumbel, Weibull ou Fréchet —, sendo, por isso, contínua.

Para o estudo da estrutura de dependência da distribuição k -variada de valores extremos G seria útil remover qualquer influência de aspetos marginais e, para esse efeito, bastaria transformar cada uma das distribuições marginais numa distribuição marginal comum. Em Resnick (1987) [38], prova-se que podemos assumir, sem perda de generalidade, que cada distribuição marginal univariada G_j , $j = 1, \dots, k$, de G é Fréchet unitária.

Proposição 3. *Sejam (Z_1, \dots, Z_k) um vetor aleatório com f.d. G , cujas distribuições marginais univariadas, G_1, \dots, G_k , são contínuas, e $(t_1(Z_1), \dots, t_k(Z_k))$ um vetor aleatório com f.d. \tilde{G} , onde*

$$t_j(\cdot) = -\frac{1}{\ln G_j(\cdot)}, \quad j \in \{1, \dots, k\}.$$

Então, as variáveis aleatórias $t_1(Z_1), \dots, t_k(Z_k)$ têm distribuição comum Fréchet unitária, e as f.d. G e \tilde{G} possuem a mesma estrutura de dependência.

Demonstração.

Sejam $\tilde{G}_1, \dots, \tilde{G}_k$ as f.d. marginais univariadas do vetor $(t_1(Z_1), \dots, t_k(Z_k))$.

Então, para cada $j = 1, \dots, k$, tem-se

$$\begin{aligned}\tilde{G}_j(z) &= P(t_j(Z_j) \leq z) \\ &= P\left(-\frac{1}{\ln G_j(Z_j)} \leq z\right) \\ &= P\left(\ln G_j(Z_j) \leq -\frac{1}{z}\right) \\ &= P\left(G_j(Z_j) \leq \exp\left(-\frac{1}{z}\right)\right) \\ &= \exp\left(-\frac{1}{z}\right), \quad z > 0,\end{aligned}$$

onde a última igualdade segue de $G_j(Z_j)$ ter distribuição uniforme, pelo Teorema da Transformação Uniformizante³. Provamos, assim, que as v.a. $t_j(Z_j)$, $j = 1, \dots, k$, têm f.d. comum Fréchet unitária.

Mostremos, agora, que as f.d. G e \tilde{G} têm a mesma estrutura de dependência.

É evidente que $G_j(\mathbb{R}) = [0, 1] = \tilde{G}_j(\mathbb{R})$, para $j \in \{1, \dots, k\}$. Além disso,

$$\begin{aligned}\tilde{G}(t_1(z_1), \dots, t_k(z_k)) &= P\left(t_1(Z_1) \leq t_1(z_1), \dots, t_k(Z_k) \leq t_k(z_k)\right) \\ &= P\left(-\frac{1}{\ln G_1(Z_1)} \leq -\frac{1}{\ln G_1(z_1)}, \dots, -\frac{1}{\ln G_k(Z_k)} \leq -\frac{1}{\ln G_k(z_k)}\right) \\ &= P(\ln G_1(Z_1) \leq \ln G_1(z_1), \dots, \ln G_k(Z_k) \leq \ln G_k(z_k)) \\ &= P(Z_1 \leq z_1, \dots, Z_k \leq z_k) \\ &= G(z_1, \dots, z_k)\end{aligned}$$

para qualquer $(z_1, \dots, z_k) > \mathbf{0}$. Pelo Lema 1, tem-se que $C_G \equiv C_{\tilde{G}}$, o que prova que as f.d. G e \tilde{G} têm a mesma estrutura de dependência. □

Mediante a suposição das distribuições marginais univariadas da distribuição k -variada de valores extremos G Fréchet unitárias, obtém-se a caracterização de G apresentada no próximo teorema (Pickands (1981) [37]). Relembre-se, do Exemplo 1, que uma possibilidade para os coeficientes de atração de uma f.d. F_j para uma Fréchet unitária são $\delta_{n,j} = n$ e $\lambda_{n,j} = 0$, onde $j \in \{1, \dots, k\}$ e $n \in \mathbb{N}$.

³**Teorema da Transformação Uniformizante:** Seja Z uma v.a. com f.d. F_Z estritamente crescente. Então $Y = F_Z(Z)$ tem distribuição uniforme $U((0, 1))$.

Teorema 3. *Seja $\{Z_n = (Z_{n,1}, \dots, Z_{n,k})\}_{n \in \mathbb{N}}$ uma sucessão de vetores aleatórios independentes com função de distribuição comum F e distribuições marginais univariadas pertencentes ao max-domínio de atração da distribuição Fréchet unitária. Se*

$$P\left(\frac{M_{n,1}}{n} \leq z_1, \dots, \frac{M_{n,k}}{n} \leq z_k\right) \xrightarrow{n \rightarrow +\infty} G(z_1, \dots, z_k), \quad \forall (z_1, \dots, z_k) \in \mathcal{C}(G),$$

então

$$G(z_1, \dots, z_k) = \exp(-V(z_1, \dots, z_k)), \quad z_j \in \mathbb{R}^+, j \in \{1, \dots, k\}, \quad (1.18)$$

sendo a função $V(\cdot)$, designada medida expoente, definida por

$$V(z_1, \dots, z_k) = \int_{S_k} \bigvee_{j=1}^k \frac{w_j}{z_j} dH(w_1, \dots, w_k),$$

com $H(\cdot)$ uma medida finita sobre o simplex $(k-1)$ -dimensional $S_k = \{y \in [0, 1]^k : y_1 + \dots + y_k = 1\}$ tal que

$$\int_{S_k} w_j dH(w_1, \dots, w_k) = 1, \quad \text{para cada } j \in \{1, \dots, k\}. \quad (1.19)$$

A medida $H(\cdot)$, designada medida espectral ou medida angular, caracteriza completamente a estrutura de dependência de uma distribuição k -variada de valores extremos (cf. Simpson (2020) [42]). Todavia, para dimensões superiores a 3, a especificação de $H(w)$, $w \in S_k$, ou, alternativamente, da medida expoente $V(z)$, $z \in \mathbb{R}_+^k$, já é difícil, o que conduziu ao aparecimento de coeficientes que quantificam a interdependência extremal. Apresentaremos alguns nas duas próximas subsecções.

A próxima proposição enuncia duas propriedades da medida expoente que vão ser úteis no restante desenvolvimento da teoria.

Proposição 4. *A medida expoente $V(\cdot)$ satisfaz as seguintes propriedades:*

(i) $V(az_1, \dots, az_k) = \frac{1}{a} V(z_1, \dots, z_k)$, para $a \in \mathbb{R}^+$ (homogénea de grau -1);

(ii) $V(+\infty, \dots, z_j, \dots, +\infty) = \frac{1}{z_j}$;

para $z_j \in \mathbb{R}^+$, $j \in \{1, \dots, k\}$.

Demonstração.

(i) Para $a \in \mathbb{R}^+$, tem-se, pela definição da medida expoente, que

$$\begin{aligned} V(az_1, \dots, az_k) &= \int_{S_k} \bigvee_{j=1}^k \frac{w_j}{az_j} dH(w_1, \dots, w_k) \\ &= \frac{1}{a} \int_{S_k} \bigvee_{j=1}^k \frac{w_j}{z_j} dH(w_1, \dots, w_k) \\ &= \frac{1}{a} V(z_1, \dots, z_k), \quad z_1, \dots, z_k \in \mathbb{R}^+. \end{aligned}$$

(ii) Pela definição de medida expoente e por (1.19), obtém-se

$$\begin{aligned}
V(+\infty, \dots, z_j, \dots, +\infty) &= \int_{S_k} \left\{ 0 \vee \dots \vee \frac{w_j}{z_j} \vee \dots \vee 0 \right\} dH(w_1, \dots, w_k) \\
&= \int_{S_k} \frac{w_j}{z_j} dH(w_1, \dots, w_k) \\
&= \frac{1}{z_j} \int_{S_k} w_j dH(w_1, \dots, w_k) \\
&= \frac{1}{z_j}, \quad z_j \in \mathbb{R}^+, j \in \{1, \dots, k\}.
\end{aligned}$$

□

A alínea (i) da proposição anterior permite provar a max-estabilidade de G nas condições do Teorema 3. Relembre-se que, em tais condições, G tem marginais univariadas Fréchet unitárias e, por isso, $\tilde{\delta}_{n,j} = n$ e $\tilde{\lambda}_{n,j} = 0$, para $j \in \{1, \dots, k\}$ e $n \in \mathbb{N}$. Assim, tem-se

$$\begin{aligned}
G^n(nz_1, \dots, nz_k) &= (\exp(-V(nz_1, \dots, nz_k)))^n \\
&= \exp(-nV(nz_1, \dots, nz_k)) \\
&= \exp(-V(z_1, \dots, z_k)) \\
&= G(z_1, \dots, z_k).
\end{aligned}$$

Ao contrário do caso univariado, onde uma família paramétrica de distribuições abrange todos os limites possíveis (recorde (1.13)), a expressão (1.18) implica que as distribuições MEV não podem ser totalmente descritas por um número finito de parâmetros. Não obstante, existem modelos paramétricos que representam subclasses das distribuições MEV. No exemplo seguinte, apresentamos um deles.

Exemplo 3. Seja $\{Z_n = (Z_{n,1}, \dots, Z_{n,k})\}_{n \in \mathbb{N}}$ uma sucessão de vetores aleatórios com função de distribuição comum

$$F(z_1, \dots, z_k) = 1 - \left(\sum_{j=1}^k z_j^{-1/\alpha} \right)^\alpha, \quad z_1, \dots, z_k \in \mathbb{R}^+,$$

para $\alpha \in (0, 1]$. Como as distribuições marginais univariadas pertencem ao max-domínio de atração da distribuição Fréchet unitária, então, tem-se

$$\begin{aligned}
\lim_{n \rightarrow +\infty} F^n(nz_1, \dots, nz_k) &= \lim_{n \rightarrow +\infty} \left[1 - \left(\sum_{j=1}^k (nz_j)^{-1/\alpha} \right)^\alpha \right]^n \\
&= \lim_{n \rightarrow +\infty} \left[1 - (n^{-1/\alpha})^\alpha \left(\sum_{j=1}^k z_j^{-1/\alpha} \right)^\alpha \right]^n
\end{aligned}$$

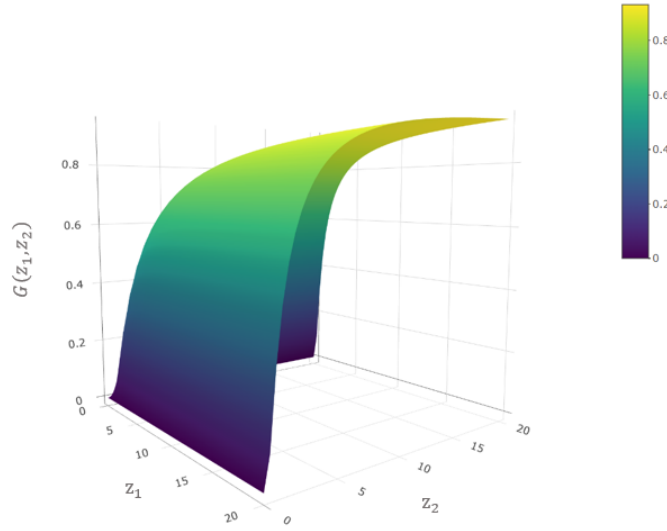


Figura 1.4: Representação gráfica da distribuição logística bivariada com $\alpha = 0.5$.

$$\begin{aligned}
 &= \lim_{n \rightarrow +\infty} \left[1 - \frac{\left(\sum_{j=1}^k z_j^{-1/\alpha} \right)^\alpha}{n} \right]^n \\
 &= \exp \left(- \left(\sum_{j=1}^k z_j^{-1/\alpha} \right)^\alpha \right) \\
 &= G(z_1, \dots, z_k).
 \end{aligned}$$

A condição (1.14) é, então, verificada, já que G tem margens univariadas não degeneradas (Fréchet unitárias). Assim, G é uma distribuição k -variada de valores extremos e é designada distribuição logística k -variada. A sua representação gráfica para $k = 2$ e $\alpha = 0.5$ encontra-se na Figura 1.4.

Além disso, por ter margens com f.d. comum Fréchet unitária, a medida expoente correspondente é dada por

$$V(z_1, \dots, z_k) = \left(\sum_{j=1}^k z_j^{-1/\alpha} \right)^\alpha,$$

atendendo ao Teorema 3. O parâmetro α indica a intensidade da dependência entre as variáveis. No caso em que $\alpha = 1$, tem-se a independência das variáveis. Quando $\alpha \rightarrow 0$, as variáveis são totalmente dependentes. Tal é visível nos gráficos de dispersão, apresentados na Figura 1.5, relativos a dados de modelos logísticos bivariados simulados no R , para três valores diferentes de α : 0.05, 0.5 e 1.

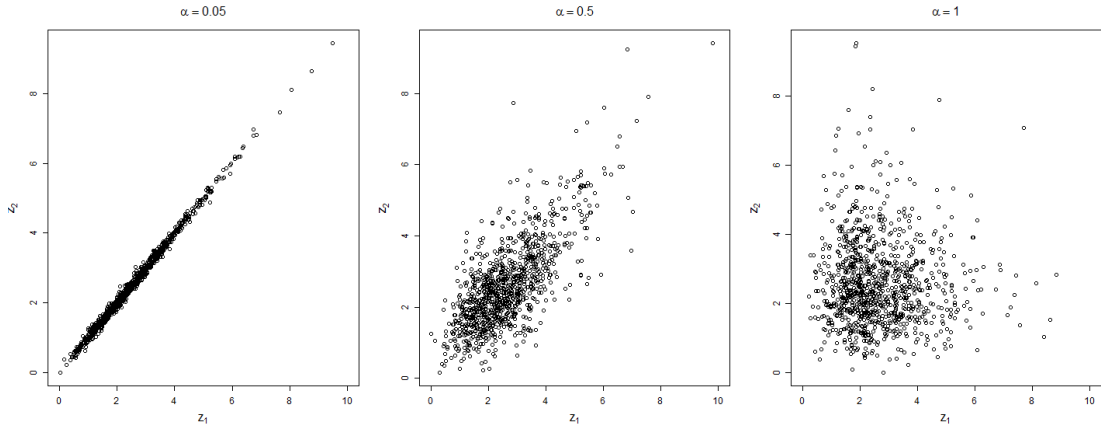


Figura 1.5: Gráficos de dispersão de dados do modelo logístico bivariado para $\alpha = 0.05$ (à esquerda), $\alpha = 0.5$ (no meio) e $\alpha = 1$ (à direita).

1.2.2 Dependência entre extremos univariados

Na extensão da TVE do contexto univariado para o multivariado, surge uma nova questão: avaliar e quantificar a dependência entre as margens de um vetor com distribuição multivariada de valores extremos, bem como a dependência entre os acontecimentos extremos que lhes estão associados. Este é um dos tópicos fundamentais na análise de valores extremos multivariados e será o tema a desenvolver de seguida. Começamos por obter um enquadramento para a f.d. MEV, onde o minorante e o majorante correspondem, respetivamente, aos casos de independência e dependência total entre as margens.

Em 1966, Lehmann [28] introduziu o conceito de *dependência no ortante positivo* entre variáveis aleatórias.

Definição 5. Seja $Z = (Z_1, \dots, Z_k)$ um vetor aleatório com função de distribuição G com distribuições marginais univariadas G_j , $j \in \{1, \dots, k\}$. As variáveis aleatórias Z_1, \dots, Z_k são **dependentes no ortante positivo** se

$$G(z_1, \dots, z_k) \geq \prod_{j=1}^k G_j(z_j), \quad (1.20)$$

e

$$P(Z_1 > z_1, \dots, Z_k > z_k) \geq \prod_{j=1}^k P(Z_j > z_j), \quad (1.21)$$

para $z_1, \dots, z_k \in \mathbb{R}$.

Intuitivamente, a expressão (1.21) significa que é mais provável que Z_1, \dots, Z_k assumam valores grandes simultaneamente do que no caso em que são independentes.

Um ano depois, Esary, Proschan e Walkup (1967) [13] apresentaram o conceito de *dependência positiva (ou associação)* entre variáveis aleatórias.

Definição 6. Seja $\mathbf{Z} = (Z_1, \dots, Z_k)$ um vetor aleatório com função de distribuição G . As variáveis aleatórias Z_1, \dots, Z_k dizem-se **associadas** (ou G diz-se uma **distribuição associada**) se, para todo o par de funções $\phi, \psi : \mathbb{R}^k \rightarrow \mathbb{R}$ não decrescentes em cada uma das suas componentes, se tem

$$\text{Cov}(\phi(\mathbf{Z}), \psi(\mathbf{Z})) \geq 0,$$

desde que $E[\phi(\mathbf{Z})], E[\psi(\mathbf{Z})], E[\phi(\mathbf{Z})\psi(\mathbf{Z})]$ existam.

Esary, Proschan e Walkup (1967) [13] mostraram ainda que este último conceito é mais forte do que o conceito de dependência no ortante positivo.

Marshall e Olkin (1983) [32] provaram que qualquer distribuição multivariada de valores extremos G é uma distribuição associada. Então, como o conceito de associação é mais forte do que o de dependência no ortante positivo, tem-se que, para um ve.a. $\mathbf{Z} = (Z_1, \dots, Z_k)$ com distribuição k -variada de valores extremos G e distribuições marginais univariadas $G_j, j \in \{1, \dots, k\}$,

$$G(z_1, \dots, z_k) \geq \prod_{j=1}^k G_j(z_j). \quad (1.22)$$

Por outro lado,

$$G(z_1, \dots, z_k) \leq G_j(z_j), \forall j \in \{1, \dots, k\}$$

donde

$$G(z_1, \dots, z_k) \leq \bigwedge_{j=1}^k G_j(z_j), \quad (1.23)$$

para quaisquer $z_1, \dots, z_k \in \mathbb{R}$.

Atendendo a (1.22) e (1.23), obtemos o seguinte enquadramento para G :

$$\prod_{j=1}^k G_j(z_j) \leq G(z_1, \dots, z_k) \leq \bigwedge_{j=1}^k G_j(z_j), \quad (1.24)$$

onde $z_j \in \mathbb{R}, j = 1, \dots, k$. Os limites inferior e superior correspondem aos casos em que as margens de $\mathbf{Z} \sim G$ são independentes e totalmente dependentes, respetivamente.

Quando as distribuições marginais univariadas de G são, em particular, Fréchet unitárias, as desigualdades em (1.24) assumem a forma

$$\bigvee_{j=1}^k \frac{1}{z_j} \leq V(z_1, \dots, z_k) \leq \sum_{j=1}^k \frac{1}{z_j}, \quad (1.25)$$

para qualquer $z_j \in \mathbb{R}^+, j = 1, \dots, k$.

De facto, atendendo ao Teorema 3,

$$\begin{aligned}
\prod_{j=1}^k G_j(z_j) &\leq G(z_1, \dots, z_k) \leq \bigwedge_{j=1}^k G_j(z_j) \\
\Leftrightarrow \prod_{j=1}^k G_j(z_j) &\leq \exp(-V(z_1, \dots, z_k)) \leq \bigwedge_{j=1}^k G_j(z_j) \\
\Leftrightarrow \ln \left(\prod_{j=1}^k G_j(z_j) \right) &\leq -V(z_1, \dots, z_k) \leq \ln \left(\bigwedge_{j=1}^k G_j(z_j) \right) \\
\Leftrightarrow \sum_{j=1}^k \ln G_j(z_j) &\leq -V(z_1, \dots, z_k) \leq \bigwedge_{j=1}^k \ln G_j(z_j) \\
\Leftrightarrow -\bigwedge_{j=1}^k \ln G_j(z_j) &\leq V(z_1, \dots, z_k) \leq -\sum_{j=1}^k \ln G_j(z_j) \\
\Leftrightarrow -\bigwedge_{j=1}^k -\frac{1}{z_j} &\leq V(z_1, \dots, z_k) \leq \sum_{j=1}^k \frac{1}{z_j} \\
\Leftrightarrow \bigvee_{j=1}^k \frac{1}{z_j} &\leq V(z_1, \dots, z_k) \leq \sum_{j=1}^k \frac{1}{z_j}.
\end{aligned}$$

Neste caso, os limites inferior e superior correspondem, respetivamente, à dependência total e independência das v.a. Z_1, \dots, Z_k .

De seguida, apresentamos dois coeficientes que permitem quantificar a dependência entre os acontecimentos $\{Z_1 \leq z_1\}, \dots, \{Z_k \leq z_k\}$, quando $z_1 = \dots = z_k = z \in \mathbb{R}$, para $k = 2$ e para $k \geq 2$, respetivamente, onde (Z_1, \dots, Z_k) é um ve.a. com distribuição k -variada de valores extremos.

Para $k = 2$, ou seja, no contexto bivariado, introduz-se o popular índice de dependência extremal de Tiago de Oliveira (1962/63) [35].

Definição 7. *Seja (Z_1, Z_2) um ve.a. com f.d. BEVG e distribuições marginais iguais a G_1 . O índice de dependência extremal de G é a constante $\varepsilon_{Z_1, Z_2} \in [1, 2]$ tal que*

$$P(Z_1 \leq z, Z_2 \leq z) = G(z, z) = G_1^{\varepsilon_{Z_1, Z_2}}(z), \quad z \in \mathbb{R}. \quad (1.26)$$

Este coeficiente é igual a um se e só se os acontecimentos $\{Z_1 \leq z\}$ e $\{Z_2 \leq z\}$ forem totalmente dependentes; e igual a dois se e somente se forem independentes.

Exemplo 4. Seja (Z_1, Z_2) um vetor aleatório com função de distribuição, G , BEV apresentada no Exemplo 2, i.e.,

$$G(z_1, z_2) = \exp[-\exp(-z_1)] \exp[-\exp(-z_2)] \exp[\exp(z_1) + \exp(z_2)]^{-1}, \quad z_1, z_2 \in \mathbb{R}^+.$$

Tendo em conta que as distribuições marginais G_1 e G_2 de G são Gumbel, podemos escrever

$$G_1(z) = G_2(z) = \Lambda_{0,1}(z) = \exp[-\exp(-z)], \quad \text{para } z \in \mathbb{R}^+.$$

Então, pela definição (1.26),

$$\begin{aligned} G(z, z) &= G_1^{\varepsilon_{Z_1, Z_2}}(z) \\ &\Leftrightarrow \exp[-2\exp(-z)] \exp[2\exp(z)]^{-1} = \exp[-\exp(-z)]^{\varepsilon_{Z_1, Z_2}} \\ &\Leftrightarrow \exp\left[-\frac{3}{2}\exp(-z)\right] = \exp[-\varepsilon_{Z_1, Z_2}\exp(-z)] \\ &\Leftrightarrow \varepsilon_{Z_1, Z_2} = \frac{3}{2}. \end{aligned}$$

Exemplo 5. Seja (Z_1, Z_2) um vetor aleatório com função de distribuição G bivariada de valores extremos apresentada no Exemplo 3, i.e.,

$$G(z_1, z_2) = \exp\left\{-\left[\left(\frac{1}{z_1}\right)^{1/\alpha} + \left(\frac{1}{z_2}\right)^{1/\alpha}\right]^\alpha\right\}, \quad z_1, z_2 \in \mathbb{R}^+, \quad \alpha \in (0, 1]. \quad (1.27)$$

Como as distribuições marginais de G , G_1 e G_2 , são Fréchet unitárias,

$$G_1(z) = G_2(z) = \exp\left(-\frac{1}{z}\right), \quad z \in \mathbb{R}^+.$$

Então,

$$G(z, z) = \exp\left(-\frac{2^\alpha}{z}\right) = G_1^{2^\alpha}(z).$$

Por conseguinte, o índice de dependência extremal de G é igual a 2^α . A sua representação gráfica encontra-se na Figura 1.6.

Para $\alpha = 1$, tem-se $\varepsilon_{Z_1, Z_2} = 2$, i.e., os acontecimentos $\{Z_1 \leq z\}$ e $\{Z_2 \leq z\}$ são independentes, para $z \in \mathbb{R}^+$. Já no caso em que $\alpha \rightarrow 0^+$, obtém-se $\varepsilon_{Z_1, Z_2} = 1$, ou seja, os acontecimentos são totalmente dependentes.

Para a quantificação da dependência entre os acontecimentos $\{Z_1 \leq z\}, \dots, \{Z_k \leq z\}$, $z \in \mathbb{R}$, com $k \geq 2$, apresenta-se o coeficiente de dependência extremal de Smith (1990) [44], que é uma generalização do índice anterior.

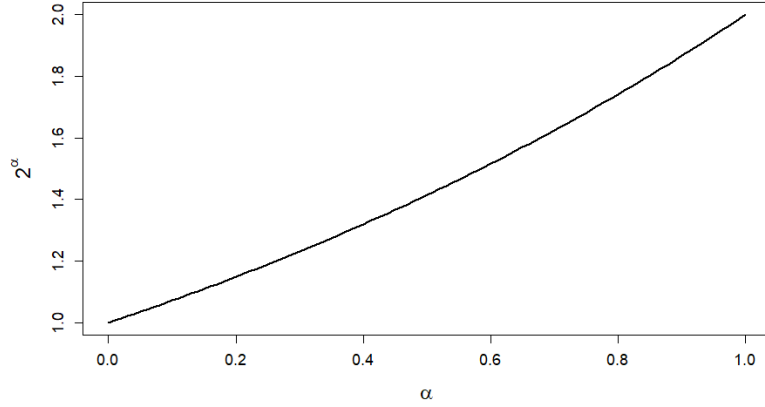


Figura 1.6: Gráfico do índice de dependência extremal da distribuição BEV logística.

Definição 8. Seja (Z_1, \dots, Z_k) um ve.a. com f.d. multivariada de valores extremos G com f.d. marginais univariadas iguais a G_1 . A constante $\varepsilon_{Z_1, \dots, Z_k} \in [1, k]$ tal que

$$G(z, \dots, z) = G_1^{\varepsilon_{Z_1, \dots, Z_k}}(z), \quad z \in \mathbb{R}, \quad (1.28)$$

designa-se coeficiente de dependência extremal.

Este coeficiente é igual a um se e somente se os acontecimentos $\{Z_1 \leq z\}, \dots, \{Z_k \leq z\}$, $z \in \mathbb{R}$, forem totalmente dependentes, e igual a k se e só se forem independentes. Além disso, quando as variáveis Z_1, \dots, Z_k são Fréchet unitárias, relaciona-se com a medida expoente do seguinte modo

$$\varepsilon_{Z_1, \dots, Z_k} = V(1, \dots, 1). \quad (1.29)$$

De facto, pelo Teorema 3,

$$\begin{aligned} G(z, \dots, z) &= \exp(-V(z, \dots, z)) \\ &= \exp\left(-\frac{1}{z}V(1, \dots, 1)\right), \end{aligned} \quad (1.30)$$

sendo que a segunda igualdade segue da alínea (i) da Proposição 4. Uma vez que as distribuições marginais de $(Z_1, \dots, Z_k) \sim G$ são iguais a G_1 , com G_1 Fréchet unitária, tem-se

$$G_1^{\varepsilon_{Z_1, \dots, Z_k}}(z) = \exp\left(-\frac{1}{z}\varepsilon_{Z_1, \dots, Z_k}\right). \quad (1.31)$$

Assim, atendendo à definição do coeficiente, e a (1.30) e (1.31), segue-se que

$$\exp\left(-\frac{1}{z}\varepsilon_{Z_1, \dots, Z_k}\right) = \exp\left(-\frac{1}{z}V(1, \dots, 1)\right),$$

donde

$$\varepsilon_{Z_1, \dots, Z_k} = V(1, \dots, 1).$$

Exemplo 6. Calculemos o coeficiente extremal de Smith para a distribuição logística apresentada no Exemplo 3. Neste, concluímos que

$$V(z_1, \dots, z_k) = \left(\sum_{j=1}^k z_j^{-1/\alpha} \right)^\alpha,$$

para $z_1, \dots, z_k \in \mathbb{R}^+$, com $\alpha \in (0, 1]$. Em particular, para $z_1 = \dots = z_k = 1$,

$$V(1, \dots, 1) = \left(\sum_{j=1}^k 1^{-1/\alpha} \right)^\alpha = k^\alpha.$$

E, pela relação (1.29),

$$\varepsilon_{Z_1, \dots, Z_k} = k^\alpha.$$

Exemplo 7. Seja $F(z_1, \dots, z_k) = 1 - \exp\left(-\bigwedge_{j=1}^k z_j\right)$, com $(z_1, \dots, z_k) \in \mathbb{R}_+^k$.

Para $(z_1, \dots, z_k) \in \mathbb{R}_+^k$, tem-se

$$\begin{aligned} F^n(z_1 + \ln n, \dots, z_k + \ln n) &= \left[1 - \exp\left(-\bigwedge_{j=1}^k (z_j + \ln n)\right) \right]^n \\ &= \left[1 - \frac{1}{n} \exp\left(-\bigwedge_{j=1}^k z_j\right) \right]^n \\ &\xrightarrow{n \rightarrow +\infty} \exp\left(-\exp\left(-\bigwedge_{j=1}^k z_j\right)\right). \end{aligned}$$

Af.d. $G(z_1, \dots, z_k) = \exp\left(-\exp\left(-\bigwedge_{j=1}^k z_j\right)\right) = \bigwedge_{j=1}^k \exp(-\exp(-z_j))$ é k -variada de valores extremos com distribuições marginais univariadas Gumbel.

Atendendo à Proposição 3, a estrutura de dependência dos vetores $(Z_1, \dots, Z_k) \sim G$ e $(t_1(Z_1), \dots, t_k(Z_k))$ é a mesma. Logo, como

$$\begin{aligned} P(t_1(Z_1) \leq z_1, \dots, t_k(Z_k) \leq z_k) &= P(Z_1 \leq -\ln(1/z_1), \dots, Z_k \leq -\ln(1/z_k)) \\ &= \bigwedge_{j=1}^k \exp(-1/z_j) \\ &= \exp\left(-\bigvee_{j=1}^k 1/z_j\right), \end{aligned}$$

a medida expoente é dada por

$$V(z_1, \dots, z_k) = \prod_{j=1}^k z_j^{-1},$$

para $z_1, \dots, z_k \in \mathbb{R}^+$.

Então, por (1.29),

$$\varepsilon_{Z_1, \dots, Z_k} = V(1, \dots, 1) = 1,$$

donde se conclui que os acontecimentos $\{Z_1 \leq z\}, \dots, \{Z_k \leq z\}$ são totalmente dependentes, para $z \in \mathbb{R}^+$.

1.2.3 Dependência entre extremos multivariados

Até aqui, centrámos a análise na dependência entre extremos univariados. No entanto, a avaliação da dependência entre extremos multivariados revela-se particularmente relevante no estudo da ocorrência de acontecimentos extremos para campos aleatórios em regiões disjuntas de localizações. Por conseguinte, estendemos agora a análise à dependência entre extremos multivariados. Iniciamos com um enquadramento teórico da função de distribuição multivariada de valores extremos, em que o minorante e o majorante correspondem, respetivamente, aos casos de independência e dependência total entre subvetores do vetor com distribuição MEV. A partir deste, obtém-se um enquadramento para a medida expoente associada, sendo ambos úteis para o desenvolvimento da teoria subsequente. Nas últimas décadas, têm sido propostas várias medidas para quantificar a dependência entre extremos multivariados (*cf.*, por exemplo, Li (2009) [30], Ferreira (2011) [14], Ferreira e Ferreira (2012) [16], Marcon et al. (2017) [31]), entre as quais destacamos uma em particular, que será introduzida no final desta subsecção.

Seja $\mathbf{Z} = (Z_1, \dots, Z_k)$ um vetor aleatório com distribuição k -variada de valores extremos G e distribuições marginais univariadas iguais a F , com F Fréchet unitária. Considere-se $\mathcal{I} = \{I_1 = \{i_0 + 1, \dots, i_1\}, I_2 = \{i_1 + 1, \dots, i_2\}, \dots, I_p = \{i_{p-1} + 1, \dots, i_p = k\}\}$ uma partição do conjunto $\{1, \dots, k\}$, com $0 = i_0 < i_1 < i_2 < \dots < i_{p-1} < i_p = k$ e $1 \leq p \leq k$. Denota-se por \mathbf{Z}_{I_j} o subvetor de \mathbf{Z} indexado em I_j , *i.e.*,

$$\mathbf{Z}_{I_j} = (Z_{i_{j-1}+1}, \dots, Z_{i_j}),$$

por $G_{\mathbf{Z}_{I_j}}$ a distribuição de valores extremos associada ao vetor \mathbf{Z}_{I_j} e por $V_{\mathbf{Z}_{I_j}}$ a correspondente medida expoente, para $j \in \{1, \dots, p\}$.

Como G é uma distribuição k -variada de valores extremos,

$$G(z_1, \dots, z_k) \geq \prod_{j=1}^p G_{\mathbf{Z}_{I_j}}(z_{i_{j-1}+1}, \dots, z_{i_j}), \quad z_1, \dots, z_k \in \mathbb{R}, \quad (1.32)$$

(*cf.* Lelowitz (1972) [29]).

Por outro lado, a desigualdade

$$G(z_1, \dots, z_k) \leq \bigwedge_{j=1}^p G_{\mathbf{Z}_{I_j}}(z_{i_{j-1}+1}, \dots, z_{i_j}), \quad (1.33)$$

também é válida para quaisquer $z_1, \dots, z_k \in \mathbb{R}$.

Assim, obtém-se o seguinte enquadramento para a distribuição k -variada de valores extremos G

$$\prod_{j=1}^p G_{\mathbf{Z}_{I_j}}(z_{i_{j-1}+1}, \dots, z_{i_j}) \leq G(z_1, \dots, z_k) \leq \bigwedge_{j=1}^p G_{\mathbf{Z}_{I_j}}(z_{i_{j-1}+1}, \dots, z_{i_j}), \quad (1.34)$$

onde $z_1, \dots, z_k \in \mathbb{R}$. O limite inferior corresponde ao caso de independência entre os subvetores $\mathbf{Z}_{I_1}, \dots, \mathbf{Z}_{I_p}$ de $\mathbf{Z} \sim G$, enquanto o superior corresponde à dependência total.

Nas condições do Teorema 3, tem-se o intervalo de valores

$$\bigvee_{j=1}^p V_{\mathbf{Z}_{I_j}}(z_{i_{j-1}+1}, \dots, z_{i_j}) \leq V_{\mathbf{Z}}(z_1, \dots, z_k) \leq \sum_{j=1}^p V_{\mathbf{Z}_{I_j}}(z_{i_{j-1}+1}, \dots, z_{i_j}), \quad (1.35)$$

para a medida expoente de $\mathbf{Z} \sim G$, para quaisquer $z_1, \dots, z_k \in \mathbb{R}^+$. Agora, os limites inferior e superior correspondem, respetivamente, aos casos de dependência total e independência.

De seguida, apresentamos uma função muito útil para o último capítulo que permite mensurar a dependência entre as margens multivariadas de uma distribuição MEV.

Função coeficiente extremal regional

No que se segue, procuraremos estender a função de dependência extremal introduzida por Ferreira e Ferreira (2018) [17].

Começamos por introduzir um lema que alicerçará não só a definição da função de dependência a apresentar, mas também propriedades relativas aos madogramas a explorar no último capítulo.

Lema 2. *Seja $\mathbf{Z} = (Z_1, \dots, Z_k)$ um vetor aleatório com distribuição k -variada de valores extremos e distribuições marginais univariadas iguais a F , com F Fréchet unitária. Considere-se $\mathcal{I} = \{I_1 = \{i_0 + 1, \dots, i_1\}, I_2 = \{i_1 + 1, \dots, i_2\}, \dots, I_p = \{i_{p-1} + 1, \dots, i_p = k\}\}$ uma partição do conjunto $\{1, \dots, k\}$, com $0 = i_0 < i_1 < i_2 < \dots < i_{p-1} < i_p = k$ e $1 \leq p \leq k$. Então*

$$P \left(\bigvee_{i \in I_1} F(Z_i) \leq u_1, \dots, \bigvee_{i \in I_p} F(Z_i) \leq u_p \right) = \exp \left(-V_{\mathbf{Z}_{\mathcal{I}}} \left(-\frac{1}{\ln u_1}, \dots, -\frac{1}{\ln u_p} \right) \right), \quad (1.36)$$

para $u_1, \dots, u_p \in (0, 1)$. Para $u_j \in (0, 1)$, com $j \in S$ e $\emptyset \neq S \subseteq \{1, \dots, p\}$, tem-se

$$P \left(\bigcap_{j \in S} \left\{ \bigvee_{i \in I_j} F(Z_i) \leq u_j \right\} \right) = \exp \left(-V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(-\sum_{j \in S} \frac{\delta_{\alpha(\bigcup_{j \in S} I_j)}(I_j)}{\ln u_j}, \dots, -\sum_{j \in S} \frac{\delta_{\omega(\bigcup_{j \in S} I_j)}(I_j)}{\ln u_j} \right) \right), \quad (1.37)$$

onde $\alpha(\bigcup_{j \in S} I_j)$ e $\omega(\bigcup_{j \in S} I_j)$ denotam, respetivamente, o primeiro e o último elementos de $\bigcup_{j \in S} I_j$ e $\delta_i(A) = 1$, se $i \in A$, e $\delta_i(A) = 0$, se $i \notin A$.

Demonstração. A distribuição conjunta de $\bigvee_{i \in I_j} F(Z_i)$ para todos os elementos I_j da partição \mathcal{I} , i.e., para I_j com $j \in \{1, \dots, p\}$, é dada por

$$\begin{aligned}
& P \left(\bigvee_{i \in I_1} F(Z_i) \leq u_1, \dots, \bigvee_{i \in I_p} F(Z_i) \leq u_p \right) \\
&= P(F(Z_1) \leq u_1, \dots, F(Z_{i_1}) \leq u_1, \dots, F(Z_{i_{p-1}+1}) \leq u_p, \dots, F(Z_k) \leq u_p) \\
&= P(Z_1 \leq F^{-1}(u_1), \dots, Z_{i_1} \leq F^{-1}(u_1), \dots, Z_{i_{p-1}+1} \leq F^{-1}(u_p), \dots, Z_k \leq F^{-1}(u_p)) \\
&= P \left(Z_1 \leq -\frac{1}{\ln u_1}, \dots, Z_{i_1} \leq -\frac{1}{\ln u_1}, \dots, Z_{i_{p-1}+1} \leq -\frac{1}{\ln u_p}, \dots, Z_k \leq -\frac{1}{\ln u_p} \right) \\
&= \exp \left(-V_{\mathbf{Z}_{\mathcal{I}}} \left(-\frac{1}{\ln u_1}, \dots, -\frac{1}{\ln u_1}, \dots, -\frac{1}{\ln u_p}, \dots, -\frac{1}{\ln u_p} \right) \right), \quad (u_1, \dots, u_p) \in (0, 1)^p,
\end{aligned}$$

uma vez que $\mathbf{Z}_{\mathcal{I}} = \mathbf{Z}$ tem distribuição k -variada de valores extremos com distribuições marginais univariadas iguais a F , com $F(z) = \exp(-z^{-1})$, para $z > 0$, e a equação (1.18) é válida. Provamos, assim, (1.36), que é equivalente à identidade

$$P \left(\bigvee_{i \in I_1} F(Z_i) \leq u_1, \dots, \bigvee_{i \in I_p} F(Z_i) \leq u_p \right) = \exp \left(-V_{\mathbf{Z}_{\mathcal{I}}} \left(-\sum_{j=1}^p \frac{\delta_1(I_j)}{\ln u_j}, \dots, -\sum_{j=1}^p \frac{\delta_k(I_j)}{\ln u_j} \right) \right), \quad (1.38)$$

atendendo à forma como a partição $\mathcal{I} = \{I_1, \dots, I_p\}$ de $\{1, \dots, k\}$ está definida (os elementos $1, \dots, k$ surgem por ordem crescente em I_1, \dots, I_p). Desta última igualdade deduz-se que a distribuição conjunta de $\bigvee_{i \in I_j} F(Z_i)$, para $j \in S$, com $\emptyset \neq S \subseteq \{1, \dots, p\}$, é dada por

$$P \left(\bigcap_{j \in S} \left\{ \bigvee_{i \in I_j} F(Z_i) \leq u_j \right\} \right) = \exp \left(-V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(-\sum_{j \in S} \frac{\delta_{\alpha(\bigcup_{j \in S} I_j)}(I_j)}{\ln u_j}, \dots, -\sum_{j \in S} \frac{\delta_{\omega(\bigcup_{j \in S} I_j)}(I_j)}{\ln u_j} \right) \right),$$

para $u_j \in (0, 1)$, sendo que $\alpha(\bigcup_{j \in S} I_j)$ e $\omega(\bigcup_{j \in S} I_j)$ representam, respetivamente, o menor e o maior elementos de $\bigcup_{j \in S} I_j$, o que prova a segunda identidade do lema. \square

Podemos agora definir a seguinte função

$$m_{\mathbf{Z}_{I_j}, j \in S}(\lambda_{I_j}, j \in S) = E \left[\bigvee_{j \in S} \bigvee_{i \in I_j} F^{\lambda_{I_j}}(Z_i) \right], \quad (1.39)$$

para os subvetores \mathbf{Z}_{I_j} do vetor $\mathbf{Z} = (Z_1, \dots, Z_k)$, onde $I_j \in \mathcal{I}$, sendo $\mathcal{I} = \{I_1, \dots, I_p\}$ uma partição do conjunto $\{1, \dots, k\}$ nas condições acima, e para $\lambda_{I_j} \in (0, +\infty)$, com $j \in S$ e $\emptyset \neq S \subseteq \{1, \dots, p\}$, que se relaciona com a medida expoente, como é estabelecido na próxima proposição.

Proposição 5. *Seja $Z = (Z_1, \dots, Z_k)$ um vetor aleatório com distribuição k -variada de valores extremos e distribuições marginais univariadas iguais a F , com F Fréchet unitária. Para $\mathcal{I} = \{I_1, \dots, I_p\}$ uma partição de $\{1, \dots, k\}$, nas condições acima, e para $\lambda_{I_j} \in (0, +\infty)$, com $j \in S$ e $\emptyset \neq S \subseteq \{1, \dots, p\}$, tem-se*

$$m_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S) = \frac{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha(\bigcup_{j \in S} I_j)}(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega(\bigcup_{j \in S} I_j)}(I_j) \right)}{1 + V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha(\bigcup_{j \in S} I_j)}(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega(\bigcup_{j \in S} I_j)}(I_j) \right)}, \quad (1.40)$$

para qualquer $\lambda_{I_j} \in (0, +\infty)$, com $j \in S$ e $\emptyset \neq S \subseteq \{1, \dots, p\}$

Demonstração. Como a função x^α , $\alpha \in (0, +\infty)$, é estritamente crescente, comuta com o máximo e segue-se que, pelo Lema 2 e pela alínea (i) da Proposição 4,

$$\begin{aligned} P \left(\bigvee_{j \in S} \bigvee_{i \in I_j} F^{\lambda_{I_j}}(Z_i) \leq u \right) &= P \left(\bigcap_{j \in S} \left\{ \bigvee_{i \in I_j} F^{\lambda_{I_j}}(Z_i) \leq u \right\} \right) \\ &= P \left(\bigcap_{j \in S} \left\{ \bigvee_{i \in I_j} F(Z_i) \leq u^{1/\lambda_{I_j}} \right\} \right) \\ &= \exp \left(-V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(-\sum_{j \in S} \frac{\delta_{\alpha(\bigcup_{j \in S} I_j)}(I_j)}{\ln u^{1/\lambda_{I_j}}}, \dots, -\sum_{j \in T} \frac{\delta_{\omega(\bigcup_{j \in S} I_j)}(I_j)}{\ln u^{1/\lambda_{I_j}}} \right) \right) \\ &= \exp \left(-V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(-\sum_{j \in S} \frac{\lambda_{I_j} \delta_{\alpha(\bigcup_{j \in S} I_j)}(I_j)}{\ln u}, \dots, -\sum_{j \in S} \frac{\lambda_{I_j} \delta_{\omega(\bigcup_{j \in S} I_j)}(I_j)}{\ln u} \right) \right) \\ &= \exp \left(\ln u V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha(\bigcup_{j \in S} I_j)}(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega(\bigcup_{j \in S} I_j)}(I_j) \right) \right) \\ &= u^{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha(\bigcup_{j \in S} I_j)}(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega(\bigcup_{j \in S} I_j)}(I_j) \right)}, \quad u \in (0, 1). \end{aligned}$$

Portanto,

$$\begin{aligned} m_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S) &= E \left[\bigvee_{j \in S} \bigvee_{i \in I_j} F^{\lambda_{I_j}}(Z_i) \right] \\ &= \int_0^1 V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha(\bigcup_{j \in S} I_j)}(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega(\bigcup_{j \in S} I_j)}(I_j) \right) \\ &\quad \times u^{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha(\bigcup_{j \in S} I_j)}(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega(\bigcup_{j \in S} I_j)}(I_j) \right) - 1} u \, du \end{aligned}$$

$$\begin{aligned}
&= \frac{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right)}{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right) + 1} \\
&\times \int_0^1 \left\{ V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right) + 1 \right\} \\
&\times u^{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right)} du \\
&= \frac{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right)}{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right) + 1} \\
&\times \left[u^{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right) + 1} \right]_0^1 \\
&= \frac{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right)}{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right) + 1}.
\end{aligned}$$

□

Escrevendo a relação enunciada na proposição anterior em ordem à medida expoente, obtém-se

$$V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right) = \frac{m_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S)}{1 - m_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S)},$$

o que sugere que o quociente

$$\frac{m_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S)}{1 - m_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S)}$$

pode ser considerado como uma medida de dependência entre as margens de $\mathbf{Z}_{\bigcup_{j \in S} I_j}$ sobre as regiões indexadas em I_j , para $j \in S$, com $\emptyset \neq S \subseteq \{1, \dots, p\}$. Ferreira e Ferreira (2018) [17] consideram, de facto, o quociente anterior como uma medida de dependência, mas apenas para o caso particular em que $S = \{1, \dots, p\}$. Aqui, definimos o quociente anterior como medida de dependência para qualquer $\emptyset \neq S \subseteq \{1, \dots, p\}$.

Definição 9. *Seja $\mathbf{Z} = (Z_1, \dots, Z_k)$ um vetor aleatório com distribuição k -variada de valores extremos e distribuições marginais univariadas iguais a F , com F Fréchet unitária. Para $\mathcal{I} = \{I_1, \dots, I_p\}$ uma partição de $\{1, \dots, k\}$, nas condições acima, e para $\lambda_{I_j} \in (0, +\infty)$, com $j \in S$ e $\emptyset \neq S \subseteq \{1, \dots, p\}$, define-se a medida de dependência extremal*

$$\varepsilon_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S) = \frac{m_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S)}{1 - m_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S)}. \quad (1.41)$$

Pelo exposto, tem-se

$$\varepsilon_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S) = V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_{I_j} \delta_{\alpha(\bigcup_{j \in S} I_j)}(I_j), \dots, \sum_{j \in S} \lambda_{I_j} \delta_{\omega(\bigcup_{j \in S} I_j)}(I_j) \right). \quad (1.42)$$

Quando a partição \mathcal{I} é composta por conjuntos singulares, *i.e.*, para $p = k$, $S = \{1, \dots, k\}$ e $\lambda_{I_1} = \dots = \lambda_{I_k} = 1$, $\varepsilon_{\mathbf{Z}_{1, \dots, Z_k}}(1, \dots, 1)$ coincide com o coeficiente extremal de Smith $\varepsilon_{\mathbf{Z}_{1, \dots, Z_k}}$ definido em (1.28). No caso particular em que $k = 2$, a medida de dependência introduzida coincide com o índice de Tiago de Oliveira $\varepsilon_{\mathbf{Z}_1, \mathbf{Z}_2}$ definido em (1.26).

Tem-se ainda

$$\begin{aligned} E \left[\prod_{j \in S} \prod_{i \in I_j} F^{\lambda_{I_j}}(Z(\mathbf{x}_i)) \right] &= m_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S) \\ &= \frac{\varepsilon_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S)}{1 + \varepsilon_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S)}. \end{aligned} \quad (1.43)$$

Em particular,

$$\begin{aligned} E \left[\prod_{i \in I_j} F^{\lambda_{I_j}}(Z(\mathbf{x}_i)) \right] &= m_{\mathbf{Z}_{I_j}}(\lambda_{I_j}) \\ &= \frac{\varepsilon_{\mathbf{Z}_{I_j}}(1)}{\lambda_{I_j} + \varepsilon_{\mathbf{Z}_{I_j}}(1)}, \end{aligned} \quad (1.44)$$

para $j \in S$, com $\emptyset \neq S \subseteq \{1, \dots, p\}$.

Além disso, tendo em conta o enquadramento (1.35) da medida expoente, a igualdade (1.42) permite estabelecer o seguinte intervalo de variação para a função coeficiente extremal regional:

$$\prod_{j \in S} \lambda_{I_j}^{-1} \varepsilon_{\mathbf{Z}_{I_j}}(1) \leq \varepsilon_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S) \leq \sum_{j \in S} \lambda_{I_j}^{-1} \varepsilon_{\mathbf{Z}_{I_j}}(1), \quad (1.45)$$

onde os limites inferior e superior correspondem, respetivamente, aos casos em que os subvetores \mathbf{Z}_{I_j} de \mathbf{Z} são totalmente dependentes e independentes. Com base neste enquadramento e em (1.41), deduz-se o seguinte intervalo de valores para a função $m_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S)$:

$$\frac{\prod_{j \in S} \lambda_{I_j}^{-1} \varepsilon_{\mathbf{Z}_{I_j, j \in S}}(1)}{1 + \prod_{j \in S} \lambda_{I_j}^{-1} \varepsilon_{\mathbf{Z}_{I_j, j \in S}}(1)} \leq m_{\mathbf{Z}_{I_j, j \in S}}(\lambda_{I_j}, j \in S) \leq \frac{\sum_{j \in S} \lambda_{I_j}^{-1} \varepsilon_{\mathbf{Z}_{I_j, j \in S}}(1)}{1 + \sum_{j \in S} \lambda_{I_j}^{-1} \varepsilon_{\mathbf{Z}_{I_j, j \in S}}(1)}, \quad (1.46)$$

onde o minorante e o majorante correspondem aos mesmos casos limite de dependência de (1.45).

Exemplo 8. Seja $Z = (Z_1, Z_2, Z_3, Z_4, Z_5)$ um vetor aleatório com distribuição G logística de valores extremos, i.e.,

$$G_Z(z_1, z_2, z_3, z_4, z_5) = \exp\left(-\left\{z_1^{-1/\alpha} + z_2^{-1/\alpha} + z_3^{-1/\alpha} + z_4^{-1/\alpha} + z_5^{-1/\alpha}\right\}^\alpha\right), \alpha \in (0, 1],$$

$$z_1, z_2, z_3, z_4, z_5 \in \mathbb{R}^+.$$

Calculemos a expressão da função coeficiente extremal regional para $\mathcal{I} = \{I_1 = \{1, 2\}, I_2 = \{3\}, I_3 = \{4, 5\}\}$, $\lambda_{I_1} = 2, \lambda_{I_2} = 1, \lambda_{I_3} = 1/2$ e $S = \{1, 3\}$.

Por (1.42),

$$\begin{aligned} \varepsilon_{Z_{I_1}, Z_{I_3}}(\lambda_{I_1}, \lambda_{I_3}) &= \varepsilon_{Z_1, Z_2, Z_4, Z_5}(2, 1/2) \\ &= V_{Z_1, Z_2, Z_4, Z_5}(2, 2, 1/2, 1/2). \end{aligned} \quad (1.47)$$

Ora,

$$\begin{aligned} G_{Z_1, Z_2, Z_4, Z_5}(z_1, z_2, z_4, z_5) &= \lim_{z_3 \rightarrow +\infty} G_Z(z_1, z_2, z_3, z_4, z_5) \\ &= \exp\left(-\left\{z_1^{-1/\alpha} + z_2^{-1/\alpha} + z_4^{-1/\alpha} + z_5^{-1/\alpha}\right\}^\alpha\right). \end{aligned}$$

Logo, pelo Teorema 3,

$$V_{Z_1, Z_2, Z_4, Z_5}(z_1, z_2, z_4, z_5) = \left\{z_1^{-1/\alpha} + z_2^{-1/\alpha} + z_4^{-1/\alpha} + z_5^{-1/\alpha}\right\}^\alpha.$$

Assim, substituindo em (1.47),

$$\begin{aligned} \varepsilon_{Z_{I_1}, Z_{I_3}}(\lambda_{I_1}, \lambda_{I_3}) &= \left\{2^{-1/\alpha} + 2^{-1/\alpha} + 2^{1/\alpha} + 2^{1/\alpha}\right\}^\alpha \\ &= 2^\alpha \left\{2^{-1/\alpha} + 2^{1/\alpha}\right\}^\alpha. \end{aligned}$$

Capítulo 2

Fundamentos de Geoestatística

A percepção de que muitos fenômenos que nos rodeiam variam no espaço esboçou a necessidade da sua descrição, modelação e previsão tendo em conta essa especificidade. Assim surgiu a Geoestatística. Apesar de algumas técnicas deste ramo da Estatística remontarem às décadas de 1930 e 1940, designadamente a Kolmogorov e Wiener, apenas na década de 1950 é que a Geoestatística foi verdadeiramente impulsionada pelo matemático e engenheiro civil francês George Matheron, que motivado pela pesquisa de Daniel Krige, um engenheiro mineiro sul africano, acerca dos dados de concentração de ouro numa mina, escreveu o *‘Traité de géostatistique appliquée’*. Desde então, a teoria viu enormes desenvolvimentos, tendo, atualmente, aplicações em diversos domínios, nomeadamente na Ecologia, Climatologia, Saúde Pública e Geologia. A Geoestatística permite-nos, então, estudar uma grande variedade de fenômenos, desde que os dados a eles relativos tenham uma natureza espacial. Estes dados podem ser vistos como a realização de um campo aleatório. Por essa razão, os campos aleatórios serão o nosso objeto de estudo. Começamos, assim, por introduzir a noção de campo aleatório e também algumas definições preliminares associadas. O estudo de campos aleatórios, especialmente com o objetivo de prever valores em diferentes localizações, seria extremamente difícil, ou até impossível, se o comportamento do campo no espaço fosse completamente anárquico. No entanto, os fenômenos observados geralmente permitem considerar algumas hipóteses de invariância no espaço, genericamente designadas por estacionariedade. A formulação de hipóteses de estacionariedade do campo é fundamental para analisar a dependência espacial, que pode ser quantificada por meio de medidas como o semivariograma, a função de covariância e a função de correlação. Por conseguinte, neste capítulo, para além de apresentarmos diferentes tipos de estacionariedade, definimos as medidas de dependência espacial referidas e exploramos algumas das suas propriedades. Damos especial destaque ao semivariograma, não apenas pela sua ampla aplicabilidade prática, mas também pelo seu papel central na modelação de estruturas espaciais. Por fim, apresentamos alguns modelos teóricos de funções de correlação válidos, que nos permitem derivar modelos de semivariogramas válidos, bem como de funções de covariância.

2.1 Fundamentos dos campos aleatórios

Nesta secção, apresentamos alguns fundamentos sobre campos aleatórios que podem ser aprofundados em Cressie (1993) [6] e Carvalho e Natário (2008) [3].

2.1.1 Definições e notação

Definição 10. Um **campo aleatório** é uma família de variáveis aleatórias reais $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ definidas sobre o mesmo espaço de probabilidade (Ω, \mathcal{F}, P) e indexadas por $\mathbf{x} \in \mathcal{D} \subseteq \mathbb{R}^d$, com $d \geq 2$.

O conjunto \mathcal{D} é chamado **espaço das localizações**.

Definição 11. Dois campos aleatórios $\{Z_1(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ e $\{Z_2(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, dizem-se **identicamente distribuídos** se, para qualquer inteiro $k \geq 1$ e quaisquer $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{D}$, os vetores $(Z_1(\mathbf{x}_1), \dots, Z_1(\mathbf{x}_k))$ e $(Z_2(\mathbf{x}_1), \dots, Z_2(\mathbf{x}_k))$ têm a mesma distribuição. Simbolicamente, escrevemos

$$(Z_1(\mathbf{x}_1), \dots, Z_1(\mathbf{x}_k)) \stackrel{d}{=} (Z_2(\mathbf{x}_1), \dots, Z_2(\mathbf{x}_k)).$$

Definição 12. Um campo aleatório $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, é de **segunda ordem** se

$$E[Z^2(\mathbf{x})] < +\infty, \forall \mathbf{x} \in \mathcal{D}.$$

Se um campo aleatório $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ é de segunda ordem, então possui valor médio,

$$m(\mathbf{x}) = E[Z(\mathbf{x})], \forall \mathbf{x} \in \mathcal{D},$$

e, pela desigualdade de Cauchy-Schwarz, existe a covariância entre $Z(\mathbf{x}_1)$ e $Z(\mathbf{x}_2)$,

$$C(\mathbf{x}_1, \mathbf{x}_2) = \text{cov}(Z(\mathbf{x}_1), Z(\mathbf{x}_2)) = E[(Z(\mathbf{x}_1) - m(\mathbf{x}_1))(Z(\mathbf{x}_2) - m(\mathbf{x}_2))], \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}.$$

A variância define-se por

$$\sigma^2(\mathbf{x}) = \text{var}[Z(\mathbf{x})] = C(\mathbf{x}, \mathbf{x})$$

e a função de correlação é dada por

$$\rho(\mathbf{x}_1, \mathbf{x}_2) = C(\mathbf{x}_1, \mathbf{x}_2) \{ \text{var}[Z(\mathbf{x}_1)] \times \text{var}[Z(\mathbf{x}_2)] \}^{-\frac{1}{2}}, \forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}.$$

Uma classe importante de campos aleatórios é a dos campos Gaussianos, uma vez que estes são modelos razoáveis para muitos fenómenos naturais, a estimação e inferência são simples e são especificados por $m(\mathbf{x})$ e $C(\mathbf{x}_1, \mathbf{x}_2)$.

Definição 13. O campo aleatório $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, diz-se **Gaussiano** se, para qualquer inteiro $k \geq 1$ e para qualquer conjunto finito de localizações $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$, o correspondente vetor aleatório $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$ tem distribuição gaussiana k -dimensional.

2.1.2 Algumas propriedades importantes

Se o comportamento de um campo aleatório ao longo de $\mathcal{D} \subseteq \mathbb{R}^d$ fosse completamente irregular, seria difícil ou mesmo impossível desenvolver estudos sobre o mesmo, designadamente estudos com vista à previsão de valores do campo em localizações não amostradas. Em geral, os fenómenos espaciais observados, nomeadamente fenómenos naturais, permitem considerar um certo número de hipóteses de invariância por translação em \mathcal{D} , genericamente designadas por estacionaridade.

Definição 14. Um campo aleatório $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, diz-se **fortemente estacionário** ou **estritamente estacionário** se, para qualquer inteiro $k > 1$, para qualquer conjunto finito de localizações $\{\mathbf{x}_j : j = 1, \dots, k\} \subset \mathcal{D}$ e qualquer vetor $\mathbf{h} \in \mathbb{R}^d$ tal que $\{\mathbf{x}_j + \mathbf{h} : j = 1, \dots, k\} \subset \mathcal{D}$, os ve.a. $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$ e $(Z(\mathbf{x}_1 + \mathbf{h}), \dots, Z(\mathbf{x}_k + \mathbf{h}))$ são identicamente distribuídos.

Em particular, a lei de $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ coincide com a lei de $\{Z(\mathbf{x} + \mathbf{r})\}_{\mathbf{x} \in \mathcal{D}}$, para \mathbf{r} arbitrariamente fixo em \mathbb{R}^d , ou seja, há invariância da lei do campo aleatório por translação em \mathcal{D} .

Outras definições de estacionariedade, menos restritivas, consideram apenas condições referentes a momentos de primeira ordem — como a esperança matemática — e momentos de segunda ordem — como a variância, covariância e variograma. Introduzimos, de seguida, a definição de campo aleatório fracamente estacionário.

Definição 15. Um campo aleatório $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, de segunda ordem, diz-se **fracamente estacionário** se

- (i) a sua esperança matemática é constante, i.e., $m(\mathbf{x}) = \mu, \forall \mathbf{x} \in \mathcal{D}$, com $\mu \in \mathbb{R}$;
- (ii) a função de covariância depende apenas do vetor de separação $\mathbf{h} \in \mathbb{R}^d$ entre as localizações, i.e.,

$$C(\mathbf{x} + \mathbf{h}, \mathbf{x}) = C(\mathbf{h}, \mathbf{0}), \quad \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D}. \quad (2.1)$$

Daqui em diante, por uma questão de simplicidade de escrita, quando o campo aleatório satisfizer (2.1), passamos a escrever $C(\mathbf{h})$ em vez de $C(\mathbf{h}, \mathbf{0})$.

No caso em que o campo aleatório é fracamente estacionário, a função de covariância verifica algumas propriedades importantes, nomeadamente as que enunciamos na próxima proposição.

Proposição 6. Num campo aleatório $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, fracamente estacionário, tem-se que

(i) a variância é constante e igual à função de covariância no vetor nulo, i.e., $\text{var}[Z(\mathbf{x})] = C(\mathbf{0})$, $\forall \mathbf{x} \in \mathcal{D}$;

(ii) a função de covariância é par, i.e., $C(\mathbf{h}) = C(-\mathbf{h})$, $\mathbf{h} \in \mathbb{R}^d$.

Demonstração.

(i) Para uma localização arbitrária $\mathbf{x} \in \mathcal{D}$,

$$C(\mathbf{0}) = C((\mathbf{x} + \mathbf{0}) - \mathbf{x}) = C(\mathbf{x} + \mathbf{0}, \mathbf{x}) = \text{var}[Z(\mathbf{x})];$$

(ii) Para uma localização arbitrária $\mathbf{x} \in \mathcal{D}$ e $\mathbf{h} \in \mathbb{R}^d$ tal que $\mathbf{x} - \mathbf{h} \in \mathcal{D}$,

$$C(-\mathbf{h}) = \text{cov}(Z(\mathbf{x} - \mathbf{h}), Z(\mathbf{x})) = \text{cov}(Z(\mathbf{x}), Z(\mathbf{x} - \mathbf{h})) = C(\mathbf{h}).$$

□

Se $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ é fracamente estacionário e $C(\mathbf{0}) \neq 0$, então

$$\rho(\mathbf{x} + \mathbf{h}, \mathbf{x}) = \frac{C(\mathbf{x} + \mathbf{h}, \mathbf{x})}{\sqrt{\text{var}[Z(\mathbf{x} + \mathbf{h})]\text{var}[Z(\mathbf{x})]}} = \frac{C(\mathbf{h})}{C(\mathbf{0})}.$$

Daqui concluímos que a correlação entre $Z(\mathbf{x} + \mathbf{h})$ e $Z(\mathbf{x})$ depende apenas do vetor de separação \mathbf{h} entre as localizações $\mathbf{x} + \mathbf{h}$ e \mathbf{x} . Escrevemos, portanto, $\rho(\mathbf{x} + \mathbf{h}, \mathbf{x}) = \rho(\mathbf{h})$.

Como

$$\rho(\mathbf{h}) = \frac{C(\mathbf{h})}{C(\mathbf{0})}, \tag{2.2}$$

a segunda imposição da definição de campo aleatório fracamente estacionário, (2.1), pode ser reescrita como $\rho(\mathbf{x} + \mathbf{h}, \mathbf{x}) = \rho(\mathbf{h})$. Para além disso, a função de correlação verifica as propriedades enunciadas na proposição seguinte.

Proposição 7. Num campo aleatório $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, fracamente estacionário, tem-se:

(i) $\rho(\mathbf{0}) = 1$;

(ii) $\rho(\mathbf{h}) = \rho(-\mathbf{h})$, $\mathbf{h} \in \mathbb{R}^d$.

Demonstração.

(i) Tendo em conta a relação (2.2),

$$\rho(\mathbf{0}) = \frac{C(\mathbf{0})}{C(\mathbf{0})} = 1.$$

(ii) Para $\mathbf{h} \in \mathbb{R}^d$, verifica-se

$$\rho(\mathbf{h}) = \frac{C(\mathbf{h})}{C(\mathbf{0})} = \frac{C(-\mathbf{h})}{C(\mathbf{0})} = \rho(-\mathbf{h}),$$

por (2.2) e pela alínea (ii) da Proposição 6.

□

Apesar de um campo aleatório fortemente estacionário de segunda ordem ser fracamente estacionário, o recíproco pode não acontecer. A título de exemplo, tem-se $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subset \mathbb{Z}^2$, constituído por v.a. independentes tais que

$$P(Z(\mathbf{x}) = t) = \begin{cases} 1 - \frac{1}{\|\mathbf{x}\|^2} & , \quad t = 0; \\ \frac{1}{2\|\mathbf{x}\|^2} & , \quad t = \|\mathbf{x}\|, \quad \forall \mathbf{x} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}; \\ \frac{1}{2\|\mathbf{x}\|^2} & , \quad t = -\|\mathbf{x}\|; \end{cases}$$

e

$$P(Z(\mathbf{0}) = t) = \begin{cases} \frac{1}{2} & , \quad t = -1; \\ \frac{1}{2} & , \quad t = 1. \end{cases}$$

Os campos Gaussianos são exemplos de campos aleatórios fracamente estacionários que também são fortemente estacionários, pois, para qualquer inteiro $k > 1$ e qualquer conjunto finito de localizações $\{\mathbf{x}_j : j = 1, \dots, k\} \subset \mathcal{D}$, a lei do ve.a. $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$ depende apenas de $E[Z(\mathbf{x}_i)]$ e de $C(\mathbf{x}_i - \mathbf{x}_j)$, $i, j \in \{1, \dots, k\}$.

Outro tipo de estacionariedade, menos exigente do que os anteriores, é o de estacionariedade intrínseca. A estacionariedade intrínseca surge com a imposição de condições explícitas não sobre as v.a. $Z(\mathbf{x})$ e $Z(\mathbf{x} + \mathbf{h})$, mas sobre as respetivas diferenças $Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})$, com $\mathbf{x} \in \mathcal{D}$ e $\mathbf{h} \in \mathbb{R}^d$ tais que $\mathbf{x} + \mathbf{h} \in \mathcal{D}$.

Definição 16. Um campo aleatório $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, diz-se **intrinsecamente estacionário** se para qualquer vetor $\mathbf{h} \in \mathbb{R}^d$, as diferenças $Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})$, com $\mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D}$, têm

(i) *esperança matemática nula, i.e., $E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 0$;*

(ii) *variância que não depende das localizações \mathbf{x} e $\mathbf{x} + \mathbf{h}$, mas sim do vetor de separação \mathbf{h} entre elas:*

$$\text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] := 2\gamma(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})) = 2\tilde{\gamma}(\mathbf{h}).^1$$

¹É comum escrever-se esta condição apenas na forma $\text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 2\gamma(\mathbf{h})$. Mas $\gamma(\cdot)$ é definido para um par de variáveis avaliadas em localizações que distam \mathbf{h} e não para a distância \mathbf{h} , i.e.,

$$2\gamma(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})) := \text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]. \quad (2.3)$$

Por isso, devemos escrever $\gamma(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})) = \tilde{\gamma}(\mathbf{h})$ quando $\gamma(\cdot)$ depende apenas de \mathbf{h} . Contudo, por simpli-

A função $2\gamma(\cdot)$ chama-se **variograma**, enquanto $\gamma(\cdot)$ designa-se **semivariograma**. Esta função desempenha um papel crucial na Geoestatística e debruçar-nos-emos sobre ela na próxima secção.

Se $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, é intrinsecamente estacionário, então

$$\text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = E[\{Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})\}^2]$$

e, conseqüentemente, (2.3) pode ser reescrita do seguinte modo:

$$2\gamma(\mathbf{h}) = E[\{Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})\}^2].$$

No restante desenvolvimento do trabalho, sempre que nos estivermos a referir à estacionariedade forte, diremos apenas que o campo aleatório é estacionário.

A noção que introduzimos de seguida — isotropia — está relacionada com a de estacionariedade no sentido de que ambas são propriedades de invariância geométrica de campos aleatórios. A estacionariedade é uma propriedade de invariância por translação em \mathcal{D} e a isotropia é uma propriedade de invariância por rotação em \mathcal{D} . Juntas produzem invariância ao movimento.

Definição 17. *Um campo aleatório $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, diz-se **isotrópico** se, para todo o $k \in \mathbb{N}$, $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{D}$, e qualquer rotação, isto é, uma matriz ortogonal $O \in \text{Isom}(\mathbb{R}^d)$ tal que $O\mathbf{x}_1, \dots, O\mathbf{x}_k \in \mathcal{D}$, os vetores $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$ e $(Z(O\mathbf{x}_1), \dots, Z(O\mathbf{x}_k))$ são identicamente distribuídos.*

Se o campo aleatório não for isotrópico, diz-se **anisotrópico**.

No caso em que o campo aleatório é isotrópico e intrinsecamente estacionário, o variograma não depende da direção do vetor de separação \mathbf{h} , mas apenas do seu comprimento, isto é, para qualquer par de v.a. $Z(\mathbf{x} + \mathbf{h})$ e $Z(\mathbf{x})$, com $\mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D}$ e $\mathbf{h} \in \mathbb{R}^d$, tem-se

$$\gamma(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})) = \gamma(\|\mathbf{h}\|),$$

onde $\|\cdot\|$ é uma norma em \mathcal{D} (usualmente a norma euclidiana). Nesse caso, o campo diz-se **homogéneo**.

Se o campo aleatório $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, é fracamente estacionário e isotrópico, a função de covariância depende apenas do comprimento do vetor de separação:

$$C(\mathbf{x} + \mathbf{h}, \mathbf{x}) = C(\mathbf{h}) = C(\|\mathbf{h}\|), \forall \mathbf{x} \in \mathcal{D}, \mathbf{h} \in \mathbb{R}^d : \mathbf{x} + \mathbf{h} \in \mathcal{D}.$$

cidade de escrita, não faremos distinção entre $\tilde{\gamma}(\cdot)$ e $\gamma(\cdot)$, escrevendo, com um certo abuso de notação,

$$2\gamma(\mathbf{h}) = \text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]$$

sempre que esta função dependa apenas de \mathbf{h} .

2.2 Funções de covariância e variogramas

De seguida, apresentamos algumas relações entre a função de covariância e o semivariograma que nos permitem estabelecer relações entre a estacionariedade intrínseca e a estacionariedade fraca, e perceber melhor a utilidade e importância do variograma.

Como

$$\begin{aligned} 2\gamma(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})) &= \text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] \\ &= \text{var}[Z(\mathbf{x} + \mathbf{h})] + \text{var}[Z(\mathbf{x})] - 2\text{cov}(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})), \end{aligned} \quad (2.4)$$

então, se o campo for fracamente estacionário, a expressão anterior pode ser reescrita na forma

$$2\gamma(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})) = 2C(\mathbf{0}) - 2C(\mathbf{h}) = 2\gamma(\mathbf{h}).$$

Isto permite-nos concluir que um campo fracamente estacionário é intrinsecamente estacionário e a igualdade

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}). \quad (2.5)$$

é válida.

Atendendo a (2.2) e (2.5), podemos relacionar a função de correlação com o semivariograma de um campo fracamente estacionário do seguinte modo:

$$\gamma(\mathbf{h}) = C(\mathbf{0})(1 - \rho(\mathbf{h})). \quad (2.6)$$

Além disso, percebemos da relação (2.5) que, conhecidas a função de covariância e a variância do campo, conseguimos determinar o semivariograma.

Será que conseguimos fazer o recíproco? Ou seja, será que conseguimos obter a função de covariância exclusivamente a partir do semivariograma para um campo intrinsecamente estacionário?

Suponhamos que o campo tem variância constante e igual a $C(\mathbf{0})$. Assim, podemos reescrever (2.4) como

$$2\gamma(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})) = 2C(\mathbf{0}) - 2\text{cov}(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})). \quad (2.7)$$

Se a condição

$$\lim_{\|\mathbf{h}\| \rightarrow +\infty} \text{cov}(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})) = 0 \quad (2.8)$$

for satisfeita, então,

$$\begin{aligned}\lim_{\|\mathbf{h}\| \rightarrow +\infty} \gamma(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})) &= \lim_{\|\mathbf{h}\| \rightarrow +\infty} \{C(\mathbf{0}) - cov(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x}))\} \\ &= C(\mathbf{0}).\end{aligned}\tag{2.9}$$

Considerando (2.7) e (2.9), tem-se ainda

$$\begin{aligned}cov(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})) &= C(\mathbf{0}) - \gamma(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})) \\ &= \lim_{\|\mathbf{u}\| \rightarrow +\infty} \gamma(Z(\mathbf{x} + \mathbf{u}), Z(\mathbf{x})) - \gamma(\mathbf{h}) \\ &= \lim_{\|\mathbf{u}\| \rightarrow +\infty} \gamma(\mathbf{u}) - \gamma(\mathbf{h})\end{aligned}\tag{2.10}$$

A partir de (2.10), obtemos uma expressão que permite determinar a covariância exclusivamente em função do semivariograma, o que é válido apenas se $\lim_{\|\mathbf{u}\| \rightarrow +\infty} \gamma(\mathbf{u})$ existir. Nesse caso, um campo intrinsecamente estacionário também será fracamente estacionário.

Por outro lado, quando o campo aleatório é apenas intrinsecamente estacionário, o semivariograma está bem definido, mas a função de covariância pode não existir. Isto evidencia que o semivariograma é uma ferramenta mais abrangente para avaliar a dependência espacial, conferindo-lhe uma importância fundamental na Geoestatística. No que se segue, apresentamos algumas das suas propriedades.

Suponhamos que estamos na presença de um campo aleatório $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, intrinsecamente estacionário. Nesse caso, o semivariograma é dado por

$$\gamma(\mathbf{h}) = \frac{1}{2} E[\{Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})\}^2],$$

para $\mathbf{x} \in \mathcal{D}$ e $\mathbf{h} \in \mathbb{R}^d$ tais que $\mathbf{x} + \mathbf{h} \in \mathcal{D}$. Daqui concluímos imediatamente que $\gamma(\mathbf{0}) \geq 0$.

Proposição 8. *Para um campo aleatório $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, intrinsecamente estacionário, tem-se*

- (i) $\gamma(\mathbf{0}) = 0$;
- (ii) $\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$.

Demonstração.

- (i) Para $\mathbf{x} \in \mathcal{D}$, tem-se $\gamma(\mathbf{0}) = \frac{1}{2} E[\{Z(\mathbf{x} + \mathbf{0}) - Z(\mathbf{x})\}^2] = 0$;
- (ii) Para $\mathbf{x} \in \mathcal{D}$ e $\mathbf{h} \in \mathbb{R}^d$ tais que $\mathbf{x} + \mathbf{h} \in \mathcal{D}$, obtém-se

$$\gamma(\mathbf{h}) = \frac{1}{2} E[\{Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})\}^2] = \frac{1}{2} E[\{Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})\}^2] = \gamma(-\mathbf{h}).$$

□

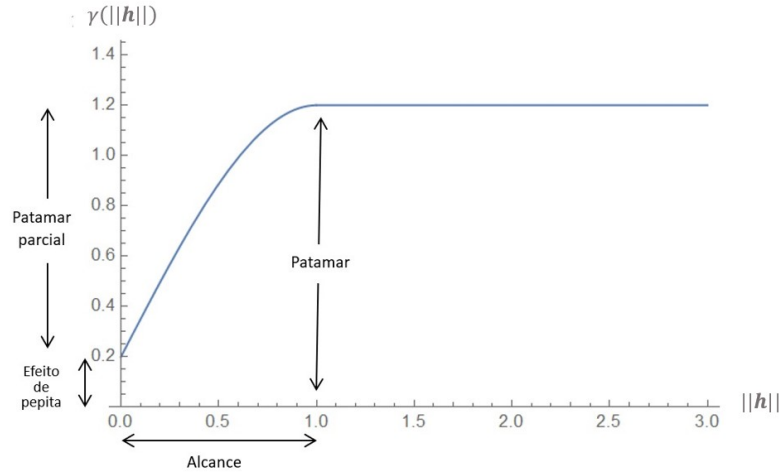


Figura 2.1: Semivariograma do modelo esférico com $\tau^2 = 0.2$ e $\sigma^2 = \phi = 1$.

Apesar de $\gamma(\mathbf{0})$ ser zero, geralmente, admite-se que o semivariograma não é contínuo na origem, *i.e.*,

$$\lim_{\|\mathbf{h}\| \rightarrow 0^+} \gamma(\mathbf{h}) = \tau^2 > 0,$$

onde o termo τ^2 é designado **efeito de pepita**.

Quando

$$\lim_{\|\mathbf{h}\| \rightarrow +\infty} \gamma(\mathbf{h}) = C(\mathbf{0}) \tag{2.11}$$

existe, é denominado **patamar**.

A diferença entre o patamar e o efeito de pepita designa-se **patamar parcial** e denota-se por σ^2 , *i.e.*,

$$\sigma^2 = C(\mathbf{0}) - \tau^2.$$

Daqui concluímos que o patamar é a soma do efeito de pepita com o patamar parcial. De facto, podemos reescrever (2.11) na forma

$$\lim_{\|\mathbf{h}\| \rightarrow +\infty} \gamma(\mathbf{h}) = \tau^2 + \sigma^2. \tag{2.12}$$

Definimos ainda o **alcance** ou **amplitude** do variograma na direção $\mathbf{h}/\|\mathbf{h}\|$ como o menor valor de $\|\mathbf{h}\|$ na direção $\mathbf{h}/\|\mathbf{h}\|$ para o qual o variograma é constante e igual ao patamar. Quando o campo subjacente é isotrópico, o alcance do variograma reduz-se ao menor valor de $\|\mathbf{h}\|$ para o qual o variograma é constante e igual ao patamar. Note-se que, para garantirmos a existência do patamar e, por sua vez, do alcance, temos de exigir que o campo subjacente seja fracamente estacionário.

Na figura 2.1, encontramos uma ilustração dos conceitos anteriores.

Vimos que o semivariograma e a função de covariância estão intimamente relacionados. As funções de covariância e de correlação devem satisfazer restrições específicas para garantir que sejam válidas, isto é, que representem corretamente a estrutura de dependência espacial de um campo aleatório. Mais especificamente, para além das propriedades já enunciadas, as funções de covariância e correlação têm de ser definidas não negativas, isto é,

$$\sum_{i=1}^k \sum_{j=1}^k a_i a_j C(\mathbf{x}_i - \mathbf{x}_j) \geq 0, \quad \sum_{i=1}^k \sum_{j=1}^k a_i a_j \rho(\mathbf{x}_i - \mathbf{x}_j) \geq 0,$$

para todo o inteiro positivo k , qualquer conjunto finito de localizações $\{\mathbf{x}_i : i = 1, \dots, k\}$ e qualquer conjunto de constantes reais $\{a_i : i = 1, \dots, k\}$. Por outro lado, o semivariograma tem de ser condicionalmente definido não positivo, ou seja,

$$\sum_{i=1}^k \sum_{j=1}^k \beta_i \beta_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \leq 0, \quad (2.13)$$

para todo o inteiro positivo k , qualquer conjunto finito de localizações $\{\mathbf{x}_i : i = 1, \dots, k\}$ e qualquer conjunto de constantes reais $\{\beta_i : i = 1, \dots, k\}$ tais que $\sum_{i=1}^k \beta_i = 0$.

Devido a estas restrições, é difícil criar modelos válidos de covariâncias ou semivariogramas. O Teorema de Bochner dá uma caracterização de funções definidas não negativas que tem sido muito utilizada para construir funções de covariâncias.

Proposição 9. *A função $C(h)$ é definida não negativa se e somente se pode ser escrita como*

$$C(h) = \int \cos(w^T h) G(dw),$$

onde a medida G , definida em \mathbb{R}^d , é limitada, positiva e simétrica em relação a 0.

Finalizamos esta secção com a apresentação de duas famílias paramétricas de funções de correlação válidas, que serão usadas no próximo capítulo, a partir das quais se podem obter famílias paramétricas de semivariogramas válidos.

A família de funções de correlação **Whittle-Matérn**, proposta por Bertil Matérn em 1960, é uma das mais populares e flexíveis para modelar a dependência espacial. A função de correlação geral é dada por

$$\rho(\|\mathbf{h}\|) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\|\mathbf{h}\|}{\phi} \right)^\nu K_\nu \left(\frac{\|\mathbf{h}\|}{\phi} \right), \quad \|\mathbf{h}\| > 0, \phi > 0,$$

onde $K_\nu(\cdot)$ é a função de Bessel modificada de ordem $\nu > 0$ e $\Gamma(\cdot)$ é a função Gama. O parâmetro ϕ é um parâmetro de escala que indica a distância a partir da qual a correlação é muito pequena. Já o parâmetro ν é um parâmetro de forma que está relacionado com a suavidade do campo: quanto maior o seu valor, mais suave é o campo, *i.e.*, menos oscilações apresenta.

Valores mais altos de ν resultam numa transição mais suave entre pontos vizinhos e num campo mais regular, enquanto valores mais baixos levam a um comportamento mais irregular e menos suave do campo. Para $\nu = 0.5$, obtém-se a função de correlação exponencial, de expressão

$$\rho(\|\mathbf{h}\|) = \exp\left(-\frac{\|\mathbf{h}\|}{\phi}\right), \|\mathbf{h}\| > 0, \phi > 0.$$

Quando $\nu \rightarrow +\infty$, a função converge para a função de correlação gaussiana

$$\rho(\|\mathbf{h}\|) = \exp\left\{-\left(\frac{\|\mathbf{h}\|}{\phi}\right)^2\right\}, \|\mathbf{h}\| > 0, \phi > 0.$$

Outra família muito conhecida é a **Exponencial Potenciada**, cuja função de correlação é dada por

$$\rho(\|\mathbf{h}\|) = \exp\left\{\left(-\frac{\|\mathbf{h}\|}{\phi}\right)^p\right\}, \|\mathbf{h}\| > 0, \phi > 0, p \in (0, 2].$$

Os parâmetros ϕ e ν são, à semelhança da família anterior, parâmetros de forma e escala, respetivamente. No entanto, enquanto ϕ continua a ter o mesmo significado, ν está agora relacionado exclusivamente com a rapidez do decaimento da correlação, sem estar diretamente associado à suavidade do campo. Por outras palavras, ν permite ajustar a forma da função de correlação, mas não está associado à continuidade ou regularidade do campo, como ocorre na Whittle-Matérn. Em particular, para $\nu = 1$, obtém-se a função de correlação exponencial, já para $\nu = 2$, obtém-se a gaussiana.

Na Figura 2.2, encontram-se as funções de correlação para as famílias paramétricas anteriores com diferentes valores de ν , considerando $\phi = 1$.

A partir das famílias de funções de correlação apresentadas, facilmente se derivam classes paramétricas de semivariogramas válidos, bem como de funções de covariância (Tabelas 2.1 e 2.2), atendendo a (2.6), (2.11) e (2.12). As suas representações gráficas para valores particulares dos parâmetros encontram-se nas Figuras 2.3 e 2.4.

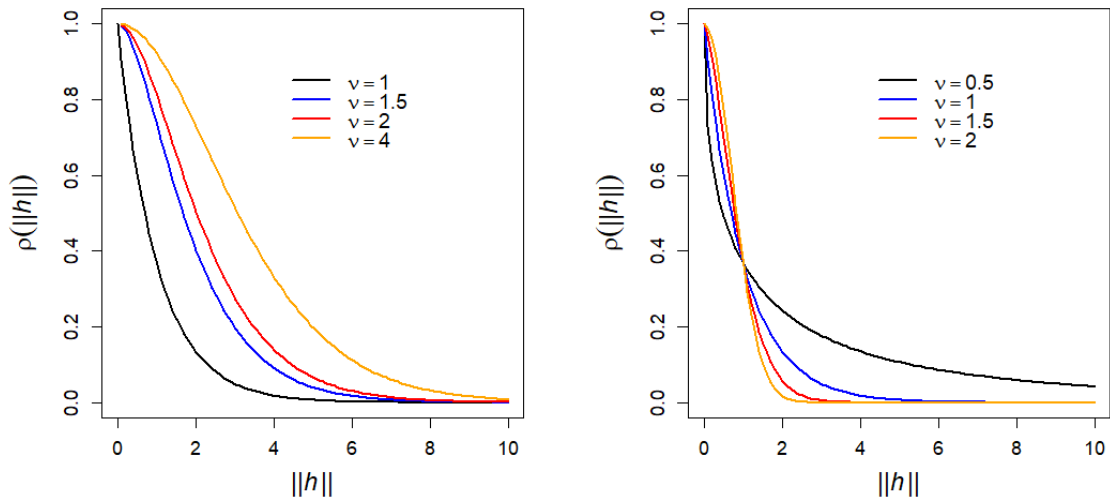


Figura 2.2: Gráficos das funções de correlação Whittle-Matérn (esquerda) e Exponencial Potenciada (direita), com $\phi = 1$, considerando $\nu \in \{0.5, 1.5, 2, 4\}$ para a Whittle-Matérn e $\nu \in \{0.5, 1, 1.5, 2\}$ para a Exponencial Potenciada.

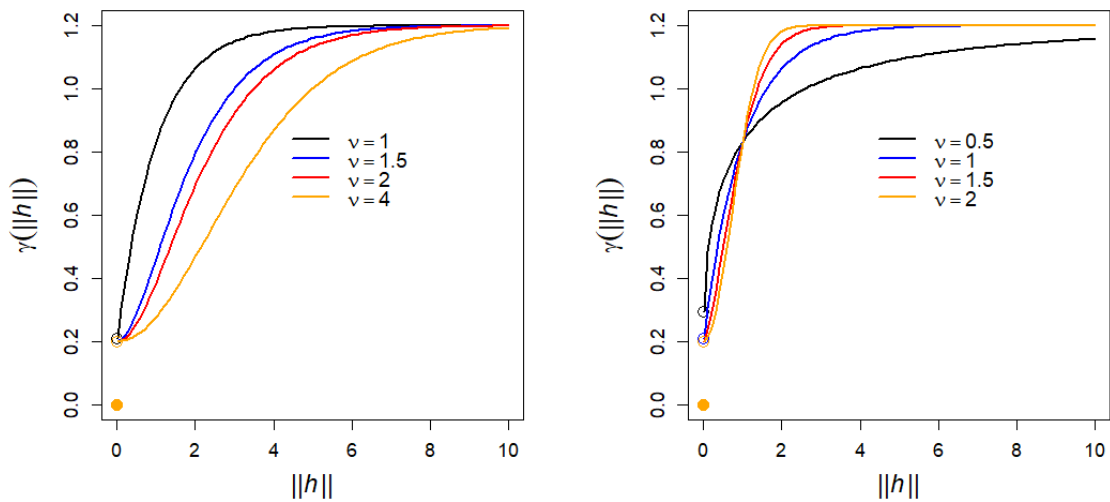


Figura 2.3: Gráficos dos semivariogramas das famílias Whittle-Matérn (esquerda) e Exponencial Potenciada (direita), com $\phi = 1 = \sigma^2, \tau^2 = 0.2$, considerando $\nu \in \{0.5, 1.5, 2, 4\}$ para a Whittle-Matérn e $\nu \in \{0.5, 1, 1.5, 2\}$ para a Exponencial Potenciada.

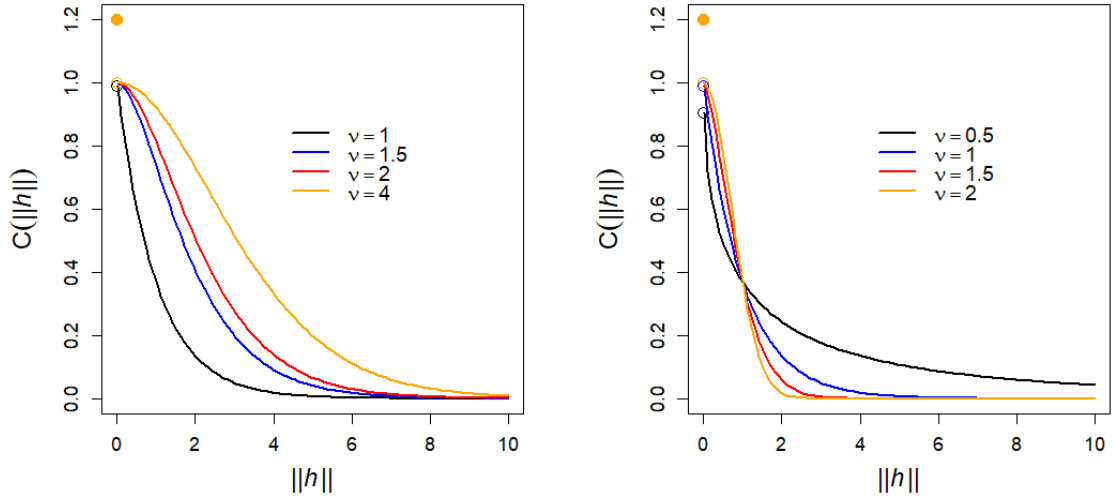


Figura 2.4: Gráficos das funções de covariância das famílias Whittle-Matérn (esquerda) e Exponencial Potenciada (direita), com $\phi = 1 = \sigma^2$, $\tau^2 = 0.2$, considerando $\nu \in \{0.5, 1.5, 2, 4\}$ para a Whittle-Matérn e $\nu \in \{0.5, 1, 1.5, 2\}$ para a Exponencial Potenciada.

Modelo	Semivariograma
Whittle-Matérn	$\gamma(\ \mathbf{h}\) = \begin{cases} \tau^2 + \sigma^2 \left(1 - \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\ \mathbf{h}\ }{\phi} \right)^\nu K_\nu \left(\frac{\ \mathbf{h}\ }{\phi} \right) \right), & \text{se } \ \mathbf{h}\ > 0; \\ 0, & \text{se } \ \mathbf{h}\ = 0. \end{cases}$ K_ν função de Bessel modificada de ordem $\nu > 0$.
Exponencial	$\gamma(\ \mathbf{h}\) = \begin{cases} \tau^2 + \sigma^2 \left(1 - \exp \left(-\frac{\ \mathbf{h}\ }{\phi} \right) \right), & \text{se } \ \mathbf{h}\ > 0, \phi > 0; \\ 0, & \text{se } \ \mathbf{h}\ = 0. \end{cases}$
Exponencial potenciado	$\gamma(\ \mathbf{h}\) = \begin{cases} \tau^2 + \sigma^2 \left(1 - \exp \left(-\left(\frac{\ \mathbf{h}\ }{\phi} \right)^p \right) \right), & \text{se } \ \mathbf{h}\ > 0, \phi > 0, p \in (0, 2]; \\ 0, & \text{se } \ \mathbf{h}\ = 0. \end{cases}$
Gaussiano	$\gamma(\ \mathbf{h}\) = \begin{cases} \tau^2 + \sigma^2 \left(1 - \exp \left(-\left(\frac{\ \mathbf{h}\ }{\phi} \right)^2 \right) \right), & \text{se } \ \mathbf{h}\ > 0, \phi > 0; \\ 0, & \text{se } \ \mathbf{h}\ = 0. \end{cases}$

Tabela 2.1: Famílias paramétricas de semivariogramas isotrópicos válidos obtidos a partir das funções de correlação definidas anteriormente.

Modelo	Função de covariância
Whittle-Matérn	$C(\ \mathbf{h}\) = \begin{cases} \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\ \mathbf{h}\ }{\phi}\right)^\nu K_\nu\left(\frac{\ \mathbf{h}\ }{\phi}\right), & \text{se } \ \mathbf{h}\ > 0; \\ \tau^2 + \sigma^2, & \text{se } \ \mathbf{h}\ = 0. \end{cases}$ <p>K_ν função de Bessel modificada de ordem $\nu > 0$.</p>
Exponencial	$C(\ \mathbf{h}\) = \begin{cases} \sigma^2 \exp\left(-\frac{\ \mathbf{h}\ }{\phi}\right), & \text{se } \ \mathbf{h}\ > 0, \phi > 0; \\ \tau^2 + \sigma^2, & \text{se } \ \mathbf{h}\ = 0. \end{cases}$
Exponencial potenciado	$C(\ \mathbf{h}\) = \begin{cases} \sigma^2 \exp\left(-\left(\frac{\ \mathbf{h}\ }{\phi}\right)^\nu\right), & \text{se } \ \mathbf{h}\ > 0, \phi > 0, \nu \in (0, 2]; \\ \tau^2 + \sigma^2, & \text{se } \ \mathbf{h}\ = 0. \end{cases}$
Gaussiano	$C(\ \mathbf{h}\) = \begin{cases} \sigma^2 \exp\left(-\left(\frac{\ \mathbf{h}\ }{\phi}\right)^2\right), & \text{se } \ \mathbf{h}\ > 0, \phi > 0; \\ \tau^2 + \sigma^2, & \text{se } \ \mathbf{h}\ = 0 \end{cases}$

Tabela 2.2: Famílias paramétricas de funções de covariância isotrópicas válidas obtidas a partir dos semivariogramas da Tabela 2.1.

Capítulo 3

Campos aleatórios max-estáveis e dependência

A Teoria de Valores Extremos, nas suas formas univariada e multivariada, pode revelar-se insuficiente para o estudo de acontecimentos extremos, especialmente em contextos onde há uma clara variabilidade espacial. Muitos acontecimentos extremos, particularmente os climatológicos, exibem uma dinâmica que se altera de forma significativa ao longo do espaço. Não só a intensidade e frequência dos acontecimentos podem variar de localização para localização, como também a ocorrência de um acontecimento extremo num determinado ponto do espaço pode influenciar ou estar relacionada com a ocorrência de acontecimentos extremos em localizações próximas. Dado que o espaço é contínuo, e não discreto, modelar acontecimentos extremos com um número finito de variáveis não é coerente com esta complexidade espacial. Assim, surge a necessidade de recorrer às ferramentas adequadas para estudar acontecimentos espacialmente distribuídos — os campos aleatórios, apresentados no capítulo anterior. Uma classe de campos aleatórios que propicia o estudo de acontecimentos extremos no contexto espacial é a dos campos max-estáveis, sendo sobre esta que nos debruçaremos neste trabalho.

Começamos este capítulo por definir tais campos e apresentar a sua caracterização espectral, de onde é possível extrair modelos paramétricos. Apresentamos dois dos mais conhecidos: o de Schlather e o de Smith. Depois, definimos várias medidas que permitem a avaliação da dependência em campos max-estáveis para um número finito de localizações, cujas definições se inspiram na definição de semivariograma, do capítulo anterior. Estudamos as suas propriedades, bem como as suas relações com as medidas de dependência definidas no Capítulo 1 e, recorrendo ao *software R*, tentamos ilustrá-las.

3.1 Campos aleatórios max-estáveis

Os campos max-estáveis constituem uma classe fundamental para a análise de valores extremos em contextos espaciais, desempenhando um papel crucial em diversas aplicações práticas, como, por exemplo, na modelação de precipitação numa dada região (Coles e Tawn (1996) [4], Buishand et al. (2008) [2], Schlather (2002) [39], Smith (1990) [44], entre ou-

tros). Tais campos fornecem uma generalização natural dos vetores com distribuições max-estáveis para o contexto espacial, pois a propriedade de max-estabilidade é preservada em todas as distribuições marginais de dimensão finita. A sua definição formal é apresentada a seguir.

Definição 18. Um campo aleatório $Z = \{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, diz-se **max-estável**, se, para Z_1, \dots, Z_n campos aleatórios i.i.d. a Z , existirem sucessões de funções reais $\{\tilde{\delta}_n(\cdot) > 0\}_{n \in \mathbb{N}}$, $\{\tilde{\lambda}_n(\cdot)\}_{n \in \mathbb{N}}$ contínuas em \mathcal{D} tais que

$$\left\{ \frac{\bigvee_{j=1}^n Z_j(\mathbf{x}) - \tilde{\lambda}_n(\mathbf{x})}{\tilde{\delta}_n(\mathbf{x})} \right\}_{\mathbf{x} \in \mathcal{D}} \stackrel{d}{=} \{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}, \quad (3.1)$$

para todo o n natural.

Se existirem sucessões de funções reais $\{\delta_n(\cdot) > 0\}_{n \in \mathbb{N}}$ e $\{\lambda_n(\cdot)\}_{n \in \mathbb{N}}$ contínuas em \mathcal{D} tais que, para Z_1, \dots, Z_n campos aleatórios i.i.d. a Z , se tem

$$\left\{ \frac{\bigvee_{j=1}^n Z_j(\mathbf{x}) - \lambda_n(\mathbf{x})}{\delta_n(\mathbf{x})} \right\}_{\mathbf{x} \in \mathcal{D}} \xrightarrow[n \rightarrow +\infty]{d} \{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}, \quad (3.2)$$

com $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ um campo aleatório com distribuições marginais univariadas não degeneradas, então $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ é max-estável (de Haan e Ferreira (2006) [9]). Note-se que $\xrightarrow[n \rightarrow +\infty]{d}$ denota a convergência em distribuição. Esta convergência justifica assim o uso dos campos max-estáveis para modelar os extremos (no caso, máximos) espaciais.

De (3.1), segue-se que as distribuições marginais univariadas e multivariadas do campo Z são de valores extremos. Doravante, assumimos que as distribuições univariadas são Fréchet unitárias.

Definição 19. Um campo max-estável $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, com distribuições marginais univariadas Fréchet unitárias designa-se campo max-estável **simples**.

Com base nessa suposição, obtém-se a caracterização do campo à custa de um processo pontual de Poisson¹ e de uma sucessão de campos aleatórios i.i.d. dada na próxima proposição (de Haan (1984) [8]).

¹Um processo pontual de Poisson em $E \subset \mathbb{R}^d$, de intensidade $\nu(\cdot)$, é uma medida pontual aleatória $N_V(\cdot) := \sum_{i \in \mathbb{N}} \delta_{V_i}(\cdot)$, com $V_i \in E$ e δ a função delta de Dirac, tal que:

- (i) $N_V(A)$ é uma v.a. com distribuição de Poisson, com esperança igual a $\nu(A)$, para todo o boreliano $A \subset E$;
- (ii) $N_V(A_1), \dots, N_V(A_k)$ são v.a. independentes, para os borelianos $A_1, \dots, A_k \subset E$ disjuntos dois a dois, qualquer que seja o k natural.

Proposição 10. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo max-estável simples. Então*

$$\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}} \stackrel{d}{=} \left\{ \bigvee_{i \in \mathbb{N}} V_i Y_i(\mathbf{x}) \right\}_{\mathbf{x} \in \mathcal{D}}, \quad (3.3)$$

onde $N_V(\cdot) := \sum_{i \in \mathbb{N}} \delta_{V_i}(\cdot)$ é um processo pontual de Poisson definido em $(0, +\infty)$, de intensidade $\nu^{-2} d\nu$, e $Y_i, i \in \mathbb{N}$, são cópias independentes de um campo aleatório $Y = \{Y(\mathbf{x})\}_{\mathbf{x} \in \mathbb{D}}$ não negativo tal que $E[Y(\mathbf{x})] = 1, \forall \mathbf{x} \in \mathcal{D}$.

De (3.3), deduz-se que as distribuições marginais de dimensão finita do campo aleatório $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ são dadas por

$$P(Z(\mathbf{x}_1) \leq z_1, \dots, Z(\mathbf{x}_k) \leq z_k) = \exp \left(-E_Y \left[\bigvee_{j=1}^k \frac{Y(\mathbf{x}_j)}{z_j} \right] \right), \quad (3.4)$$

para $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{D}$, $z_1, \dots, z_k \in \mathbb{R}^+$ e $k \in \mathbb{N}$. A prova encontra-se na demonstração da Proposição 23 em Apêndice.

Como o vetor $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$ tem distribuição k -variada de valores extremos com margens Fréchet unitárias, por (1.18) do Teorema 3, tem-se que

$$P(Z(\mathbf{x}_1) \leq z_1, \dots, Z(\mathbf{x}_k) \leq z_k) = \exp \left(-V_{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)}(z_1, \dots, z_k) \right). \quad (3.5)$$

Logo, a medida expoente é dada por

$$V_{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)}(z_1, \dots, z_k) = E_Y \left[\bigvee_{j=1}^k \frac{Y(\mathbf{x}_j)}{z_j} \right]. \quad (3.6)$$

A representação de um campo aleatório max-estável simples $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ dada em (3.3) permitiu, assim, obter expressões para a f.d. conjunta das variáveis $Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)$ e correspondente medida expoente — que caracteriza a estrutura de dependência das mesmas —, à custa dos campos aleatórios $\{Y_i(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}, i \in \mathbb{N}$. A representação (3.3) indica a existência de infinitos campos max-estáveis. Na prática, o desafio consiste em desenvolver modelos que sejam simultaneamente flexíveis e parcimoniosos, capazes de capturar uma vasta gama de dependências extremas. A parcimônia é crucial, uma vez que os dados extremos são, geralmente, escassos, enquanto a flexibilidade é essencial para evitar ajustes inadequados, que podem levar a estimativas errôneas do risco. Têm surgido vários modelos na literatura. De seguida, apresentam-se dois dos mais conhecidos.

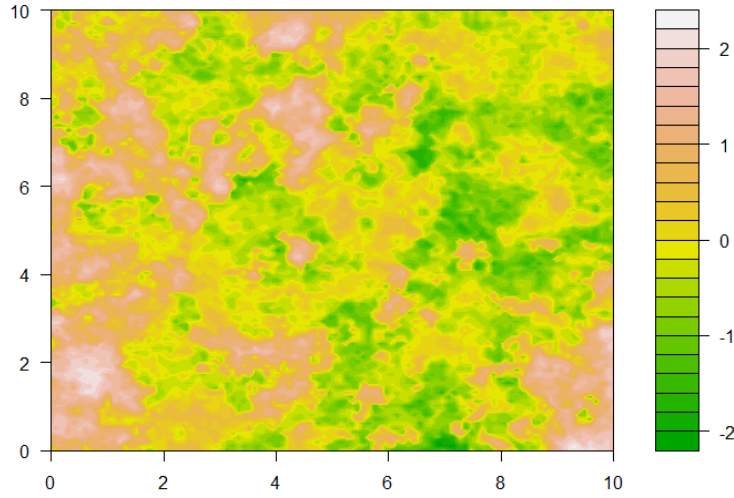


Figura 3.1: Simulação de modelo de Schlather com espaço das localizações $\mathcal{D} = [0, 10]^2$ e função de correlação exponencial para $\phi = 1$. Os valores do campo estão apresentados em escala logarítmica natural, de forma a realçar as suas variações.

1. Modelo de Schlather

O modelo de Schlather (2002) [39] obtém-se de (3.3), considerando

$$Y_i(\mathbf{x}) = \sqrt{2\pi} \{0 \vee \varepsilon_i(\mathbf{x})\}, \quad \mathbf{x} \in \mathcal{D},$$

onde ε_i são cópias independentes de um campo Gaussiano com esperança nula, variância unitária e função de correlação $\rho(\cdot)$.

A função de distribuição bivariada define-se por

$$P(Z(\mathbf{x}_1) \leq z_1, Z(\mathbf{x}_2) \leq z_2) = \exp \left[-\frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left(1 + \sqrt{1 - \frac{2\{1 + \rho(\mathbf{x}_1 - \mathbf{x}_2)\}z_1z_2}{(z_1 + z_2)^2}} \right) \right]$$

e a medida expoente é dada por

$$V_{Z(\mathbf{x}_1), Z(\mathbf{x}_2)}(z_1, z_2) = \frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left(1 + \sqrt{1 - \frac{2\{1 + \rho(\mathbf{x}_1 - \mathbf{x}_2)\}z_1z_2}{(z_1 + z_2)^2}} \right), \quad (3.7)$$

para $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ e $z_1, z_2 \in \mathbb{R}^+$ (cf. Padoan et al. (2010) [36]).

Apresentamos na Figura 3.1 uma simulação de um campo de Schlather com função de correlação da família exponencial, obtida a partir da implementação da função *rmaxstab* da biblioteca *SpatialExtremes* do R.

2. Modelo de Smith

Seja $N_U(\cdot) := \sum_{i \in \mathbb{N}} \delta_{U_i}(\cdot)$ um processo pontual de Poisson definido em \mathcal{D} , de intensidade du . Então, o modelo de Smith (1990) [44] obtém-se de (3.3), considerando

$$Y_i(\mathbf{x}) = \phi_k(\mathbf{x} - U_i; \mathbf{0}, \Sigma), \quad \mathbf{x} \in \mathcal{D},$$

onde $\phi_k(\cdot; \mathbf{0}, \Sigma)$ denota a função de densidade de probabilidade Gaussiana k -variada com esperança $\mathbf{0}$ e matriz de covariâncias Σ .

A função de distribuição bivariada é dada por

$$\begin{aligned} & P(Z(\mathbf{x}_1) \leq z_1, Z(\mathbf{x}_2) \leq z_2) \\ &= \exp \left[-\frac{1}{z_1} \Phi \left(\frac{a(\mathbf{x}_1 - \mathbf{x}_2)}{2} + \frac{1}{a(\mathbf{x}_1 - \mathbf{x}_2)} \ln \frac{z_2}{z_1} \right) - \frac{1}{z_2} \Phi \left(\frac{a(\mathbf{x}_1 - \mathbf{x}_2)}{2} + \frac{1}{a(\mathbf{x}_1 - \mathbf{x}_2)} \ln \frac{z_1}{z_2} \right) \right] \end{aligned}$$

e a medida expoente tem expressão

$$V_{Z(\mathbf{x}_1), Z(\mathbf{x}_2)}(z_1, z_2) = \frac{1}{z_1} \Phi \left(\frac{a(\mathbf{x}_1 - \mathbf{x}_2)}{2} + \frac{1}{a(\mathbf{x}_1 - \mathbf{x}_2)} \ln \frac{z_2}{z_1} \right) + \frac{1}{z_2} \Phi \left(\frac{a(\mathbf{x}_1 - \mathbf{x}_2)}{2} + \frac{1}{a(\mathbf{x}_1 - \mathbf{x}_2)} \ln \frac{z_1}{z_2} \right), \quad (3.8)$$

para $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ e $z_1, z_2 \in \mathbb{R}^+$ (cf. Padoan et al. (2010) [36]). Note-se que Φ representa a f.d. de uma v.a. Gaussiana padrão e $a(\mathbf{x}_1 - \mathbf{x}_2)$ representa a distância de Mahalanobis entre \mathbf{x}_1 e \mathbf{x}_2 , i.e.,

$$a(\mathbf{x}_1 - \mathbf{x}_2) = \sqrt{(\mathbf{x}_1 - \mathbf{x}_2)^T \Sigma^{-1} (\mathbf{x}_1 - \mathbf{x}_2)}.$$

Na Figura 3.2, representa-se graficamente uma simulação do modelo de Smith com matriz de covariâncias $\Sigma = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}$. Esta também foi obtida recorrendo à função *rmaxstab* da biblioteca *SpatialExtremes* do R.

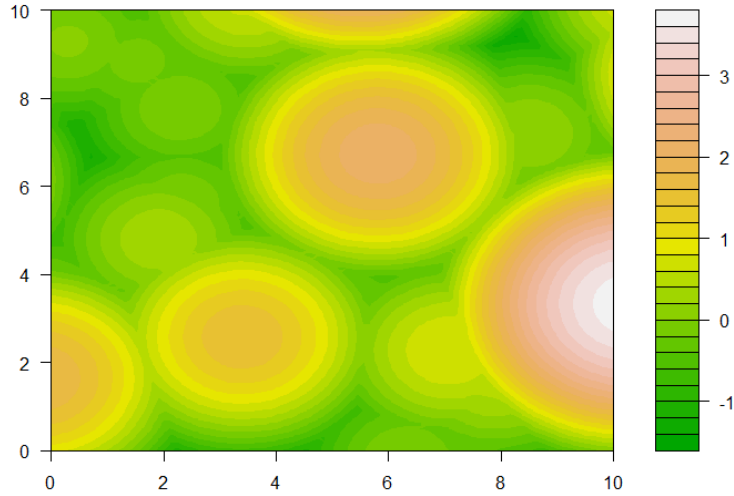


Figura 3.2: Simulação de modelo de Smith com espaço das localizações $\mathcal{D} = [0, 10]^2$ e matriz de covariâncias $\Sigma = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}$. Os valores do campo estão apresentados em escala logarítmica natural, de forma a realçar as suas variações.

3.2 Dependência

Nos estudos de extremos espaciais, a análise da estrutura de dependência do campo max-estável é fundamental para compreender como a ocorrência de valores extremos em determinadas localizações afeta a ocorrência de valores extremos em regiões vizinhas. Na prática, contudo, nem sempre é viável adotar modelos com uma estrutura de dependência previamente conhecida, tornando essencial o uso de medidas que permitam a sua avaliação, como as que veremos de seguida.

3.2.1 Função coeficiente extremal

Para a quantificação da dependência bivariada em campos max-estáveis estacionários, Schlather e Tawn (2003) [40] introduziram a função coeficiente extremal bidimensional, que é uma extensão para o contexto espacial do índice de dependência extremal de Tiago de Oliveira (1962/63) [35].

Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo estacionário max-estável simples. Para qualquer par de localizações $(\mathbf{x} + \mathbf{h}, \mathbf{x}) \in \mathcal{D} \times \mathcal{D}$, a função coeficiente extremal $\varepsilon(\mathbf{h})$, definida sobre \mathbb{R}^d e com valores em $[1, 2]$, é tal que

$$P(Z(\mathbf{x} + \mathbf{h}) \leq z, Z(\mathbf{x}) \leq z) = \exp\left(-\frac{1}{z}\varepsilon(\mathbf{h})\right), \quad (3.9)$$

para $z \in \mathbb{R}^+$.

De (3.5) sabe-se que

$$P(Z(\mathbf{x} + \mathbf{h}) \leq z, Z(\mathbf{x}) \leq z) = \exp(-V_{Z(\mathbf{x}+\mathbf{h}),Z(\mathbf{x})}(z, z)), \quad z \in \mathbb{R}^+.$$

Como a medida expoente $V_{Z(\mathbf{x}+\mathbf{h}),Z(\mathbf{x})}$ é homogénea de grau -1 (alínea (i) da Proposição 4), a equação anterior é equivalente a

$$P(Z(\mathbf{x} + \mathbf{h}) \leq z, Z(\mathbf{x}) \leq z) = \exp\left(-\frac{1}{z}V_{Z(\mathbf{x}+\mathbf{h}),Z(\mathbf{x})}(1, 1)\right), \quad z \in \mathbb{R}^+. \quad (3.10)$$

Considerando (3.9) e (3.10), obtém-se uma fórmula explícita para a função coeficiente extremal à custa da medida expoente,

$$\varepsilon(\mathbf{h}) = V_{Z(\mathbf{x}+\mathbf{h}),Z(\mathbf{x})}(1, 1). \quad (3.11)$$

Atendendo a (3.9), tem-se

$$P(Z(\mathbf{x} + \mathbf{h}) \leq z, Z(\mathbf{x}) \leq z) = P^{\varepsilon(\mathbf{h})}(Z(\mathbf{x}) \leq z), \quad z \in \mathbb{R}^+, \quad (3.12)$$

já que o campo $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ é simples. Esta relação evidencia que $\varepsilon(\mathbf{h}) \in [1, 2]$ é uma medida de dependência entre o par de acontecimentos $\{Z(\mathbf{x} + \mathbf{h}) \leq z\}$ e $\{Z(\mathbf{x}) \leq z\}$, sendo igual a um se e só se estes forem totalmente dependentes e igual a dois se e só se forem independentes. Assim, $\varepsilon(\mathbf{h})$ caracteriza parcialmente a estrutura de dependência bivariada de um campo estacionário max-estável simples.

No caso em que o campo $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ é max-estável simples, estacionário e também isotrópico, para qualquer par de localizações $(\mathbf{x} + \mathbf{h}, \mathbf{x}) \in \mathcal{D} \times \mathcal{D}$, a função coeficiente extremal está definida sobre \mathbb{R}^+ , dependendo apenas do comprimento do vetor \mathbf{h} , $\|\mathbf{h}\|$, que separa as localizações. Nesse caso, escreve-se $\varepsilon(\|\mathbf{h}\|)$, em vez de $\varepsilon(\mathbf{h})$, com um certo abuso de notação.

No exemplo seguinte, apresenta-se a expressão da função coeficiente extremal para os dois modelos de campos max-estáveis simples apresentados na secção anterior.

Exemplo 9. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ o campo de Schlather definido anteriormente com função de correlação isotrópica. Então, a função coeficiente extremal é dada por*

$$\varepsilon(\|\mathbf{h}\|) = 1 + \sqrt{\frac{1 - \rho(\|\mathbf{h}\|)}{2}}, \quad \text{para } \mathbf{h} \in \mathcal{D}.$$

De facto, usando (3.11) e (3.7), obtém-se que

$$\begin{aligned} \varepsilon(\|\mathbf{h}\|) &= V_{Z(\mathbf{x}+\mathbf{h}),Z(\mathbf{x})}(1, 1) \\ &= \frac{1}{2} \times 2 \times \left(1 + \sqrt{1 - \frac{2\{1 + \rho(\|\mathbf{h}\|)\}}{4}}\right) \\ &= 1 + \sqrt{\frac{1 - \rho(\|\mathbf{h}\|)}{2}}. \end{aligned}$$

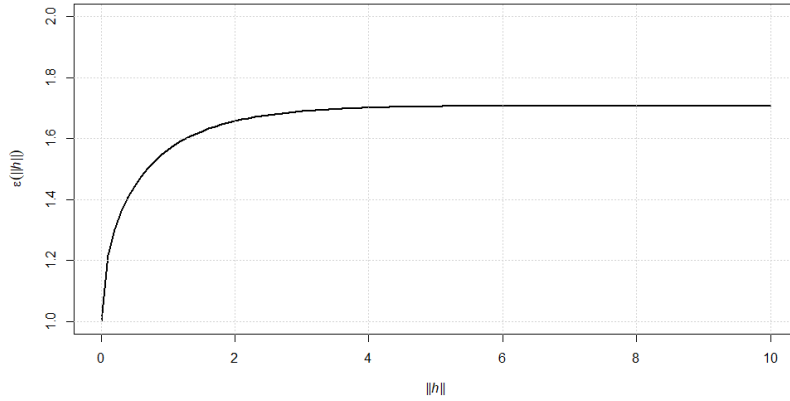


Figura 3.3: Função coeficiente extremal para modelo de Schlather isotrópico e estacionário com função de correlação exponencial com $\phi = 1$.

Se $\|\mathbf{h}\| \rightarrow 0^+$, então $\rho(\|\mathbf{h}\|) \rightarrow 1$ e, conseqüentemente, $\varepsilon(\|\mathbf{h}\|) \rightarrow 1$. Já se $\|\mathbf{h}\| \rightarrow +\infty$, é comum que $\rho(\|\mathbf{h}\|) \rightarrow 0$, o que faz com que $\varepsilon(\|\mathbf{h}\|) \rightarrow 1 + \sqrt{1/2} \simeq 1.707 \neq 2$. Isto prova que o modelo de Schlather não engloba o caso de independência. O exposto é evidenciado na Figura 3.3, onde se apresenta o gráfico da função coeficiente extremal mencionada, para $\rho(\cdot)$ pertencente ao modelo exponencial com $\phi = 1$.

No caso do modelo de Smith, obtém-se

$$\varepsilon(a(\mathbf{h})) = 2\Phi\left(\frac{a(\mathbf{h})}{2}\right), \quad \mathbf{h} = \mathbf{x}_1 - \mathbf{x}_2 \in \mathcal{D}, \quad \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D},$$

com $a(\mathbf{h}) = \sqrt{\mathbf{h}^T \Sigma^{-1} \mathbf{h}}$ e Φ a f.d. de uma v.a. Gaussiana padrão. Apresentamos um caso particular na Figura 3.4.

Os casos $a(\mathbf{h}) \rightarrow 0^+$ e $a(\mathbf{h}) \rightarrow +\infty$ correspondem, respectivamente, aos casos de dependência total e independência entre os acontecimentos $\{Z(\mathbf{x}_1) \leq z\}$ e $\{Z(\mathbf{x}_2) \leq z\}$, para $\mathbf{h} = \mathbf{x}_1 - \mathbf{x}_2$, com $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$ e $z \in \mathbb{R}^+$.

3.2.2 Madogramas

Tal como referimos no capítulo anterior, o semivariograma

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]$$

é uma ferramenta que desempenha um papel importante na avaliação da dependência espacial de um campo aleatório estacionário $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ na Geoestatística. Especificamente, o semivariograma permite analisar como a dependência entre as variáveis aleatórias $Z(\mathbf{x} + \mathbf{h})$ e $Z(\mathbf{x})$ evolui com a distância \mathbf{h} , com base na variabilidade observada entre $Z(\mathbf{x} + \mathbf{h})$ e $Z(\mathbf{x})$. Assim, ao quantificar a variabilidade entre estas variáveis, o semivariograma reflete a dependência entre elas: quanto maior a dependência, menor a variabilidade entre $Z(\mathbf{x} + \mathbf{h})$ e

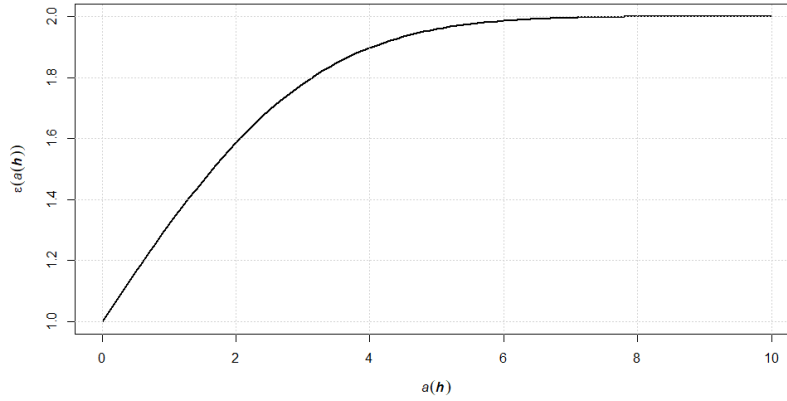


Figura 3.4: Função coeficiente extremal de um campo de Smith isotrópico com matriz de covariâncias $\Sigma = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}$.

$Z(\mathbf{x})$, o que resulta em valores mais baixos do semivariograma; por outro lado, valores mais elevados do semivariograma refletem uma variabilidade maior, o que, por sua vez, indica uma dependência mais fraca. Quando se consideram campos max-estáveis, a aplicação do semivariograma para avaliar a dependência bivariada torna-se inviável, uma vez que estes campos podem apresentar variâncias (e até esperanças) infinitas. Contudo, existem ferramentas alternativas — semivariogramas de primeira ordem ou madogramas —, cujas definições são inspiradas na do semivariograma, que possibilitam a quantificação da dependência em campos max-estáveis.

Madograma de Matheron

Em 1987, Matheron [33] introduziu o primeiro madograma, substituindo a variância da diferença entre $Z(\mathbf{x} + \mathbf{h})$ e $Z(\mathbf{x})$, que aparece na definição do semivariograma, pela esperança do valor absoluto dessa diferença. Embora não tenha sido inicialmente concebido para campos max-estáveis, o madograma ofereceu uma alternativa viável para campos com variância infinita e esperança finita. De seguida, apresentamos a sua definição formal.

Definição 20. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório intrinsecamente estacionário. O madograma é definido por*

$$\nu(\mathbf{h}) = \frac{1}{2} E[|Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})|], \quad (3.13)$$

para $\mathbf{x} \in \mathcal{D}$ e $\mathbf{h} \in \mathbb{R}^d$ tais que $\mathbf{x} + \mathbf{h} \in \mathcal{D}$.

Quando o campo aleatório é max-estável, estacionário e tem margens GEV de parâmetros (λ, δ, ξ) com ξ inferior a um, o madograma está relacionado com a função coeficiente extremal definida na Secção 3.2.1, da forma apresentada na seguinte proposição (Cooley et al., 2006 [5]):

Proposição 11. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo estacionário max-estável com distribuições marginais univariadas GEV de parâmetros (λ, δ, ξ) , $\lambda \in \mathbb{R}$, $\delta \in \mathbb{R}^+$ e $\xi \in \mathbb{R}$, e função coeficiente extremal $\varepsilon(\cdot)$. Se o índice de cauda ξ satisfizer a condição $\xi < 1$, então o madograma $\nu(\cdot)$ e a função coeficiente extremal $\varepsilon(\cdot)$ verificam*

$$\varepsilon(\mathbf{h}) = \begin{cases} u_\beta \left(\lambda + \frac{\nu(\mathbf{h})}{\Gamma(1-\xi)} \right), & \text{se } \xi \neq 0; \\ \exp \left(\frac{\nu(\mathbf{h})}{\delta} \right), & \text{se } \xi = 0; \end{cases} \quad (3.14)$$

para $\mathbf{h} \in \mathbb{R}^d$, onde $u_\beta(u) = \left(1 + \xi \frac{u-\lambda}{\delta}\right)_+^{1/\xi}$, com $a_+ = a$ se $a > 0$ e $a_+ = 0$ caso contrário, e $\Gamma(\cdot)$ é a função Gama.

Demonstração.

Atendendo à igualdade $|a - b| = 2 \{a \vee b\} - (a + b)$, para $\mathbf{x} \in \mathcal{D}$, $\mathbf{h} \in \mathbb{R}^d$ e $\mathbf{x} + \mathbf{h} \in \mathcal{D}$, tem-se

$$\begin{aligned} \nu(\mathbf{h}) &= E[Z(\mathbf{x} + \mathbf{h}) \vee Z(\mathbf{x})] - \frac{1}{2}E[Z(\mathbf{x} + \mathbf{h})] - \frac{1}{2}E[Z(\mathbf{x})] \\ &= E[Z(\mathbf{x} + \mathbf{h}) \vee Z(\mathbf{x})] - E[Z(\mathbf{x})], \end{aligned} \quad (3.15)$$

onde a última desigualdade segue do facto de $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ ser estacionário.

Como a função coeficiente extremal é invariante sob transformação das margens para a distribuição GEV (μ, σ, ξ) (cf. Schlather e Tawn (2003) [40]), tem-se

$$P(\{Z(\mathbf{x} + \mathbf{h}) \vee Z(\mathbf{x})\} \leq z) = F^{\varepsilon(\mathbf{h})}(z),$$

onde F é a f.d. marginal univariada do campo.

Então,

$$\begin{aligned} E[Z(\mathbf{x} + \mathbf{h}) \vee Z(\mathbf{x})] &= \int z \varepsilon(\mathbf{h}) F^{\varepsilon(\mathbf{h})-1}(z) dF(z) \\ &= \varepsilon(\mathbf{h}) E[Z(\mathbf{x}) F^{\varepsilon(\mathbf{h})-1}(Z(\mathbf{x}))]. \end{aligned} \quad (3.16)$$

Observe que a expressão acima pode ser interpretada em termos dos momentos ponderados de probabilidade, introduzidos por Greenwood et al. (1979) [23]. Para uma v.a. Y com f.d. F , estes são definidos por

$$M_{p,r,s} = E[Y^p [F^r(Y)][1 - F(Y)]^s], \quad p, r, s \in \mathbb{R}.$$

Assim, $E[Z(\mathbf{x} + \mathbf{h}) \vee Z(\mathbf{x})]$ pode ser escrita à custa do momento ponderado de probabilidade

$$M_{1, \varepsilon(\mathbf{h})-1, 0} = E \left[Z(\mathbf{x}) F^{\varepsilon(\mathbf{h})-1}(Z(\mathbf{x})) \right].$$

Tendo em conta que, por hipótese, F é uma GEV de parâmetros (λ, δ, ξ) , segundo Hosking et al. (1985) [24], tem-se

$$\begin{aligned} M_{1, r, 0} &= E[Z(\mathbf{x})F^r(Z(\mathbf{x}))] \\ &= \frac{1}{r+1} \left(\lambda - \frac{\delta}{\xi} + \frac{\delta}{\xi}(1+r)^\xi \Gamma(1-\xi) \right), \end{aligned}$$

para $r \geq 0$ e $0 \neq \xi < 1$. Donde se deduz que

$$E[Z(\mathbf{x})] = \lambda - \frac{\delta}{\xi} + \frac{\delta}{\xi}\Gamma(1-\xi)$$

e que

$$E\left[Z(\mathbf{x})F^{\varepsilon(\mathbf{h})-1}(Z(\mathbf{x}))\right] = \frac{1}{\varepsilon(\mathbf{h})} \left(\lambda - \frac{\delta}{\xi} + \frac{\delta}{\xi}\varepsilon^\xi(\mathbf{h})\Gamma(1-\xi) \right),$$

para $\xi < 1$ e $\xi \neq 0$. Substituindo em (3.16) e depois em (3.15), vem que

$$\nu(\mathbf{h}) = \frac{\delta}{\xi}\varepsilon^\xi(\mathbf{h})\Gamma(1-\xi) - \frac{\delta}{\xi}\Gamma(1-\xi),$$

o que é equivalente a ter

$$\varepsilon(\mathbf{h}) = \left(1 + \frac{\xi}{\delta} \frac{\nu(\mathbf{h})}{\Gamma(1-\xi)} \right)^{1/\xi} = u_\beta \left(\lambda + \frac{\nu(\mathbf{h})}{\Gamma(1-\xi)} \right).$$

Provámos, assim, (3.14) para o caso em que $\xi < 1$ e $\xi \neq 0$.

A expressão (3.14) para o caso em que $\xi = 0$ obtém-se considerando $\lim_{\xi \rightarrow 0} \varepsilon(\mathbf{h})$. De facto,

$$\begin{aligned} \lim_{\xi \rightarrow 0} \varepsilon(\mathbf{h}) &= \lim_{\xi \rightarrow 0} \left(1 + \frac{\xi}{\delta} \frac{\nu(\mathbf{h})}{\Gamma(1-\xi)} \right)^{1/\xi} \\ &= \lim_{1/\xi=y \rightarrow \infty} \left(1 + \frac{\frac{\nu(\mathbf{h})}{\delta\Gamma(1-\frac{1}{y})}}{y} \right)^y \\ &= \exp \left(\frac{\nu(\mathbf{h})}{\delta} \right). \end{aligned}$$

□

A relação (3.14) entre o madograma e a função coeficiente extremal evidencia que o madograma pode ser utilizado como medida da dependência bivariada de um campo max-estável, sob a validade das condições da Proposição 11. Esta relação permite determinar o intervalo de valores do madograma, bem como identificar os casos limite de dependência correspondentes aos extremos do intervalo, conforme estabelecido na próxima proposição.

Proposição 12. Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório max-estável estacionário com distribuições marginais univariadas GEV de parâmetros (λ, δ, ξ) , $\lambda \in \mathbb{R}$, $\delta \in \mathbb{R}^+$ e $1 > \xi \in \mathbb{R}$.

(i) Se $\xi = 0$, então $\nu(\mathbf{h}) \in [0, \delta \ln 2]$;

(ii) Se $\xi \neq 0$, então $\nu(\mathbf{h}) \in \left[0, \frac{(2^\xi - 1)\delta\Gamma(1-\xi)}{\xi}\right]$.

Os limites inferior e superior dos intervalos correspondem, respetivamente, aos casos em que os acontecimentos $\{Z(\mathbf{x} + \mathbf{h}) \leq z\}$ e $\{Z(\mathbf{x}) \leq z\}$ são, respetivamente, totalmente dependentes e independentes, para $\mathbf{x} \in \mathcal{D}$, $\mathbf{h} \in \mathbb{R}^d$, $\mathbf{x} + \mathbf{h} \in \mathcal{D}$ e $1 + \xi \frac{z-\lambda}{\delta} \in \mathbb{R}^+$.

Demonstração. De (3.14), obtém-se

$$\nu(\mathbf{h}) = \begin{cases} \frac{\delta}{\xi} \varepsilon^\xi(\mathbf{h}) \Gamma(1-\xi) - \frac{\delta}{\xi} \Gamma(1-\xi), & \text{se } \xi \neq 0; \\ \delta \ln \varepsilon(\mathbf{h}), & \text{se } \xi = 0. \end{cases}$$

O resultado segue-se tendo em conta que

$$1 \leq \varepsilon(\mathbf{h}) \leq 2,$$

onde o limite inferior corresponde ao caso de total dependência entre os acontecimentos $\{Z(\mathbf{x} + \mathbf{h}) \leq z\}$ e $\{Z(\mathbf{x}) \leq z\}$ e o limite superior ao caso de independência. □

Estimação do madograma

Um estimador natural do madograma é dado pela seguinte expressão

$$\hat{\nu}(\mathbf{h}) = \frac{1}{2T} \sum_{t=1}^T |Z^{(t)}(\mathbf{x} + \mathbf{h}) - Z^{(t)}(\mathbf{x})|, \quad (3.17)$$

onde $Z^{(t)}(\mathbf{x} + \mathbf{h})$ e $Z^{(t)}(\mathbf{x})$, $t \in \{1, \dots, T\}$, são réplicas independentes das variáveis $Z(\mathbf{x} + \mathbf{h})$ e $Z(\mathbf{x})$. No caso em que o campo é max-estável, estacionário e tem margens univariadas GEV de parâmetros (λ, δ, ξ) com $\xi < 1$, o estimador anterior permite obter o seguinte estimador para a função coeficiente extremal

$$\hat{\varepsilon}(\mathbf{h}) = \begin{cases} u_\beta \left(\lambda + \frac{\hat{\nu}(\mathbf{h})}{\Gamma(1-\xi)} \right), & \text{se } \xi \neq 0; \\ \exp \left(\frac{\hat{\nu}(\mathbf{h})}{\delta} \right), & \text{se } \xi = 0; \end{cases} \quad (3.18)$$

tendo em conta a relação dada em (3.14).

Estes estimadores estão implementados na biblioteca *SpatialExtremes* do software *R*. No exemplo seguinte, utilizá-los-emos para calcular as estimativas do madograma e da função coeficiente extremal para um conjunto de dados também disponível na biblioteca referida.

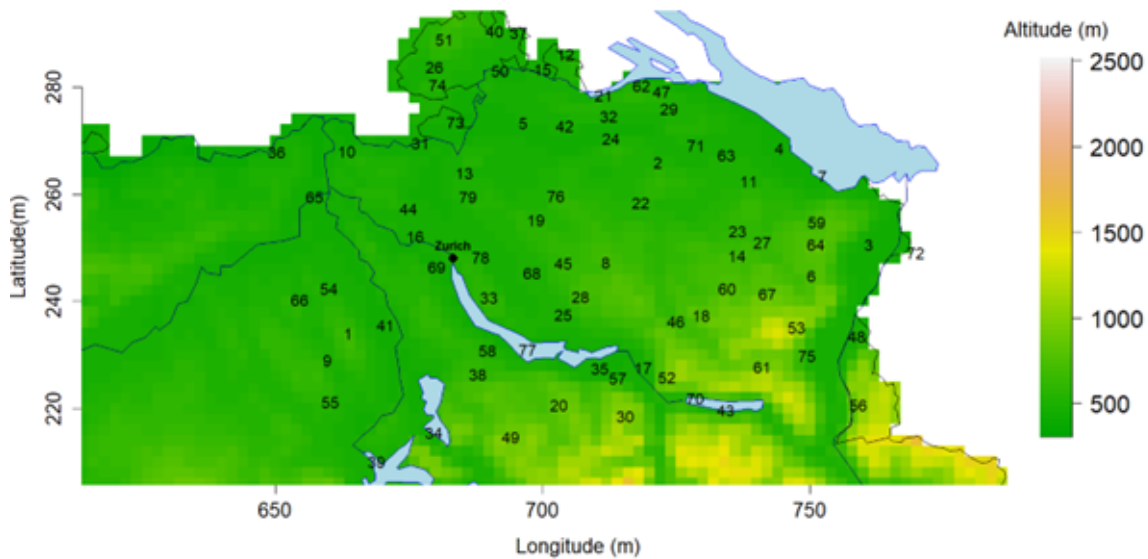


Figura 3.5: Mapa da Suíça com as localizações l_i , com índices $i \in \{1, \dots, 79\}$ assinalados no mapa, das estações de medição da precipitação do conjunto de dados *rainfall*.

Exemplo 10. *Ilustraremos agora a aplicação dos estimadores do madograma e da função coeficiente extremal, utilizando o conjunto de dados rainfall, disponível na biblioteca SpatialExtremes do software R.*

Este conjunto de dados contém 47 observações da precipitação máxima diária (em mm) para os meses de junho a agosto, entre os anos de 1962 e 2008, em 79 localizações na Suíça identificadas no mapa da Figura 3.5.

*Através do comando **madogram**, foram obtidas as estimativas para o madograma e para a função coeficiente extremal apresentadas na Figura 3.6. Estas evidenciam uma dependência espacial significativa para níveis de precipitação elevada em localizações próximas. No gráfico à esquerda, as estimativas do madograma próximas de zero para distâncias reduzidas indicam uma forte dependência bivariada para estações próximas, enquanto o seu aumento gradual com a distância sugere uma diminuição progressiva da dependência à medida que a distância aumenta.*

De forma consistente, no gráfico à direita, a função coeficiente extremal apresenta valores próximos de um para distâncias curtas, refletindo elevada dependência entre níveis de precipitação elevada para estações próximas, e aproxima-se de dois com o aumento da distância, indiciando uma transição para independência entre a precipitação elevada nas estações mais distantes.

*Note-se que, nestes cálculos, a função **madogram** transformou os dados de forma a se obterem margens Gumbel padrão ($\lambda = 0, \delta = 1$ e $\xi = 0$), o que explica a variação das estimativas do madograma entre 0 (no caso de dependência total) e $\ln 2$ (independência), cf. Proposição 12.*

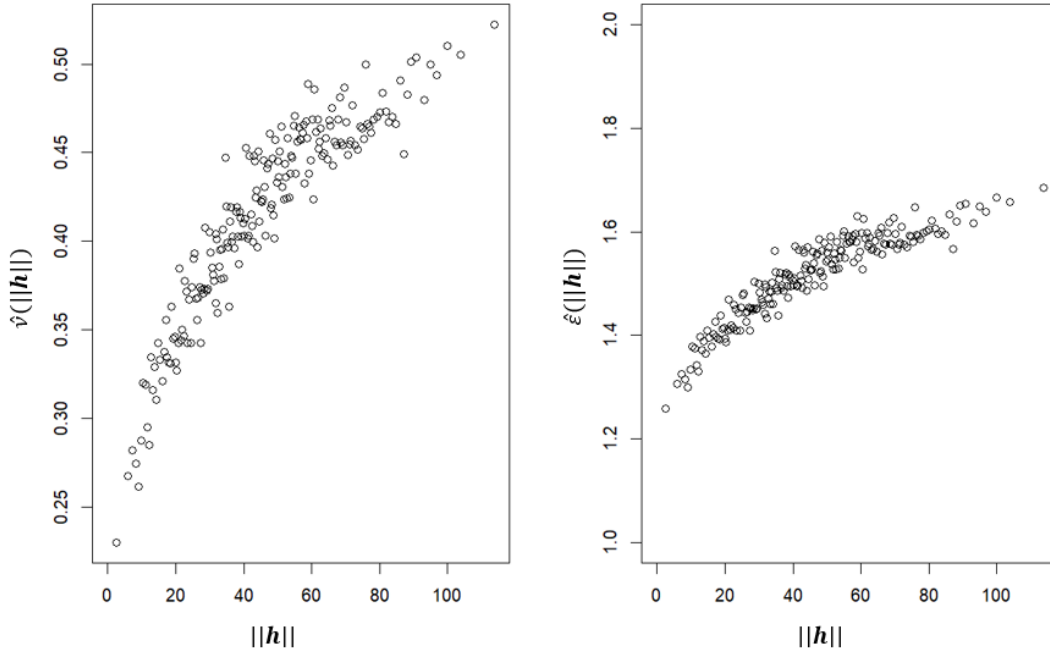


Figura 3.6: Estimativas obtidas para o madograma (gráfico à esquerda) e para a função coeficiente extremal (gráfico à direita) do conjunto de dados *rainfall*.

Madograma normalizado

Na avaliação da dependência bivariada em campos max-estáveis simples, a utilização do madograma não é viável, pois este não é finito. Para ultrapassar essa limitação, Cooley et al. (2006) [5] introduziram o madograma normalizado, também designado F -madograma, quando se assume que as distribuições marginais univariadas do campo são iguais a F . Apresentamos, de seguida, a sua definição formal, considerando, em particular, que $F(z) = \exp(-1/z)$, para $z > 0$.

Definição 21. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório max-estável estacionário com f.d. marginal univariada igual a F , com F Fréchet unitária. O madograma normalizado, também designado F -madograma, define-se por*

$$\nu_F(\mathbf{h}) = \frac{1}{2} E [|F(Z(\mathbf{x} + \mathbf{h})) - F(Z(\mathbf{x}))|], \quad (3.19)$$

para $\mathbf{x} \in \mathcal{D}$ e $\mathbf{h} \in \mathbb{R}^d$ tais que $\mathbf{x} + \mathbf{h} \in \mathcal{D}$.

Na próxima proposição, encontramos uma relação entre o madograma normalizado e a função coeficiente extremal.

Proposição 13. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório max-estável estacionário com f.d. marginal univariada igual a F , com F Fréchet unitária. Então, tem-se*

$$\nu_F(\mathbf{h}) = \frac{\varepsilon(\mathbf{h})}{\varepsilon(\mathbf{h}) + 1} - \frac{1}{2}, \quad (3.20)$$

ou equivalentemente,

$$\varepsilon(\mathbf{h}) = \frac{1 + 2\nu_F(\mathbf{h})}{1 - 2\nu_F(\mathbf{h})}, \quad (3.21)$$

para $\mathbf{h} \in \mathbb{R}^d$.

Demonstração. Relembrando a igualdade $|a - b| = 2\{a \vee b\} - (a + b)$ e tendo em conta que $E[F(Z(\mathbf{x}))] = \frac{1}{2}$, já que $F(Z(\mathbf{x})) \sim \mathcal{U}((0, 1))$, pelo Teorema da Transformação Uniformizante (3), tem-se

$$\begin{aligned} \nu_F(\mathbf{h}) &= \frac{1}{2}E[2\{F(Z(\mathbf{x} + \mathbf{h})) \vee F(Z(\mathbf{x}))\} - (F(Z(\mathbf{x} + \mathbf{h})) + F(Z(\mathbf{x})))] \\ &= E[F(Z(\mathbf{x} + \mathbf{h})) \vee F(Z(\mathbf{x}))] - \frac{1}{2} \end{aligned} \quad (3.22)$$

para $\mathbf{x} \in \mathcal{D}$ e $\mathbf{h} \in \mathbb{R}^d$ tais que $\mathbf{x} + \mathbf{h} \in \mathcal{D}$.

Considerando a partição $\mathcal{I} = \{I_1 = \{1\}, I_2 = \{2\}\}$ do conjunto $\{1, 2\}$, $\lambda_{I_1} = 1 = \lambda_{I_2}$, $S = \{1, 2\}$ e $\mathbf{x}_1 = \mathbf{x} + \mathbf{h}$ e $\mathbf{x}_2 = \mathbf{x}$, segue-se, de (1.43), que

$$E[F(Z(\mathbf{x} + \mathbf{h})) \vee F(Z(\mathbf{x}))] = \frac{\varepsilon_{Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})}(1, 1)}{1 + \varepsilon_{Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})}(1, 1)}.$$

Mas, por (1.42), $\varepsilon_{Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})}(1, 1) = V_{Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})}(1, 1)$, e, por (3.11), $V_{Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})}(1, 1) = \varepsilon(\mathbf{h})$, já que o campo $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ é estacionário. Assim,

$$E[F(Z(\mathbf{x} + \mathbf{h})) \vee F(Z(\mathbf{x}))] = \frac{\varepsilon(\mathbf{h})}{1 + \varepsilon(\mathbf{h})}. \quad (3.23)$$

Atendendo a (3.23) e (3.22), tem-se

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

ou equivalentemente

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

□

Proposição 14. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório max-estável estacionário com f.d. marginal univariada igual a F , com F Fréchet unitária. Então,*

$$\nu_F(\mathbf{h}) \in [0, 1/6],$$

sendo que o limite inferior corresponde à total dependência entre os acontecimentos $\{Z(\mathbf{x} + \mathbf{h}) \leq z\}$ e $\{Z(\mathbf{x}) \leq z\}$, enquanto o superior corresponde à independência, para $z \in \mathbb{R}^+$, $\mathbf{x} \in \mathcal{D}$ e $\mathbf{h} \in \mathbb{R}^d$ tais que $\mathbf{x} + \mathbf{h} \in \mathcal{D}$.

Demonstração. Da relação (3.20), vem que

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

Atendendo a

$$1 \leq \varepsilon(\mathbf{h}) \leq 2,$$

onde os limites inferior e superior correspondem, respetivamente, aos casos em que os acontecimentos $\{Z(\mathbf{x} + \mathbf{h}) \leq z\}$ e $\{Z(\mathbf{x}) \leq z\}$ são totalmente dependentes e independentes, obtém-se o pretendido. \square

Estimação do madograma normalizado

Seja $(Z^{(t)}(\mathbf{x} + \mathbf{h}), Z^{(t)}(\mathbf{x}))$, $t \in \{1, \dots, T\}$, uma amostra aleatória de $(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x}))$ com f.d. marginais, $F_{Z(\mathbf{x} + \mathbf{h})}$ e $F_{Z(\mathbf{x})}$, conhecidas, que poderão ser transformadas em Fréchet unitárias considerando

$$-\frac{1}{\ln F_{Z(\mathbf{x} + \mathbf{h})}(Z^{(t)}(\mathbf{x} + \mathbf{h}))} \text{ e } -\frac{1}{\ln F_{Z(\mathbf{x})}(Z^{(t)}(\mathbf{x}))}. \quad (3.24)$$

Nas condições anteriores, Cooley et al. (2006) [5] definem o seguinte estimador para o madograma normalizado

$$\hat{\nu}_F(\mathbf{h}) = \frac{1}{2T} \sum_{t=1}^T \left| F_{Z(\mathbf{x} + \mathbf{h})}(Z^{(t)}(\mathbf{x} + \mathbf{h})) - F_{Z(\mathbf{x})}(Z^{(t)}(\mathbf{x})) \right|. \quad (3.25)$$

Se $F_{Z(\mathbf{x} + \mathbf{h})}$ e $F_{Z(\mathbf{x})}$ forem desconhecidas, podem ser substituídas por estimadores $\hat{F}_{Z(\mathbf{x} + \mathbf{h})}$ e $\hat{F}_{Z(\mathbf{x})}$. Para obter estimativas mais precisas, é comum utilizar variantes da função de distribuição empírica, como

$$\hat{F}_{Z(\mathbf{y})}(u) = \frac{1}{T + 1} \sum_{t=1}^T \mathbb{1}_{\{Z^{(t)}(\mathbf{y}) \leq u\}}, \quad u \in (0, +\infty). \quad (3.26)$$

Assim sendo, obtém-se

$$\hat{\nu}_F(\mathbf{h}) = \frac{1}{2T} \sum_{t=1}^T \left| \hat{F}_{Z(\mathbf{x} + \mathbf{h})}(Z^{(t)}(\mathbf{x} + \mathbf{h})) - \hat{F}_{Z(\mathbf{x})}(Z^{(t)}(\mathbf{x})) \right|.$$

Ambos os estimadores são assintoticamente normais e fortemente consistentes.

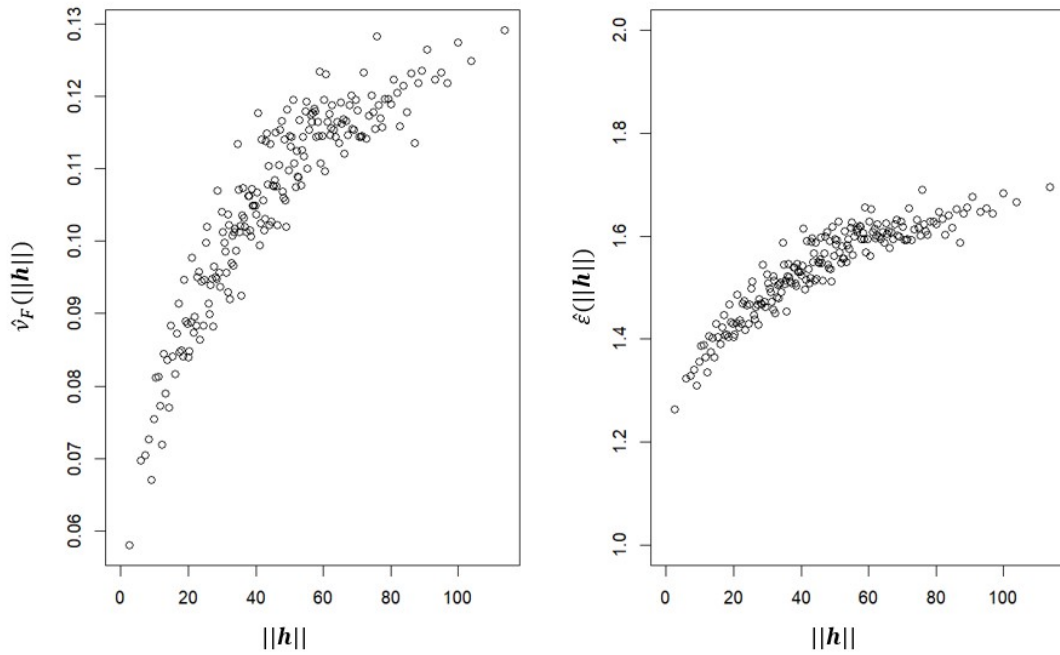


Figura 3.7: Estimativas obtidas para o F -madograma (gráfico à esquerda) e para a função coeficiente extremal (gráfico à direita) do conjunto de dados *rainfall*.

Atendendo à relação (3.21), obtém-se o seguinte estimador para $\varepsilon(\mathbf{h})$:

$$\hat{\varepsilon}(\mathbf{h}) = \frac{1 + 2\hat{\nu}_F(\mathbf{h})}{1 - 2\hat{\nu}_F(\mathbf{h})}. \quad (3.27)$$

O estimador $\hat{\varepsilon}(\mathbf{h})$ é fortemente consistente. Contudo, o facto de consistir no quociente de estimadores assintoticamente normais não permite, por si só, inferir que também seja assintoticamente normal.

Exemplo 11. Recorrendo ao conjunto de dados apresentado no Exemplo 10 e ao comando **fmadogram** da biblioteca *SpatialExtremes* do *R*, obtiveram-se as estimativas para o F -madograma e para a função coeficiente extremal apresentadas na Figura 3.7. À semelhança do observado no Exemplo 10, verifica-se que a dependência entre níveis elevados de precipitação em pares de estações diminui à medida que a distância entre estas aumenta. Contudo, as estimativas do madograma aproximam-se agora de $1/6$, em vez de $\ln 2$, tal como seria expectável pela Proposição 14.

λ -madograma

Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório estacionário max-estável com f.d. marginal univariada igual a F , com F Fréchet unitária.

De (3.12), sabe-se que

$$P(Z(\mathbf{x} + \mathbf{h}) \leq z, Z(\mathbf{x}) \leq z) = F^{\varepsilon(\mathbf{h})}(z), \quad z \in \mathbb{R}^+,$$

o que significa que $\varepsilon(\mathbf{h})$ e, conseqüentemente, $\nu_F(\mathbf{h})$ e $\nu(\mathbf{h})$ (pela relações (3.20) e (3.14)) podem ser utilizados para quantificar a dependência entre os acontecimentos $\{Z(\mathbf{x} + \mathbf{h}) \leq z_1\}$ e $\{Z(\mathbf{x}) \leq z_2\}$, mas apenas no caso particular em que $z_1 = z_2$, com $z_1, z_2 \in \mathbb{R}^+$, $\mathbf{x} \in \mathcal{D}$ e $\mathbf{h} \in \mathbb{R}^d$ tais que $\mathbf{x} + \mathbf{h} \in \mathcal{D}$. Assim, $\varepsilon(\mathbf{h})$, $\nu_F(\mathbf{h})$ e $\nu(\mathbf{h})$ não caracterizam completamente a estrutura de dependência bivariada do campo $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$.

Nos casos em que $z_1 \neq z_2$, considerando $\lambda = z_1/(z_1 + z_2)$ tem-se

$$\begin{aligned} P(Z(\mathbf{x} + \mathbf{h}) \leq z_1, Z(\mathbf{x}) \leq z_2) &= P(Z(\mathbf{x} + \mathbf{h}) \leq \lambda(z_1 + z_2), Z(\mathbf{x}) \leq (1 - \lambda)(z_1 + z_2)) \\ &= \exp(-V_{Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})}(\lambda(z_1 + z_2), (1 - \lambda)(z_1 + z_2))) \\ &= \exp\left(-\frac{1}{z_1 + z_2} V_{Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})}(\lambda, 1 - \lambda)\right) \\ &= F^{V_{Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})}(\lambda, 1 - \lambda)}(z_1 + z_2), \end{aligned}$$

por (1.18) do Teorema 3 e pela alínea (i) da Proposição 4, já que o campo é max-estável e tem distribuições marginais univariadas iguais a F , com $F(z) = \exp(-\frac{1}{z})$, para $z > 0$.

Considerando a partição $\mathcal{I} = \{I_1 = \{1\}, I_2 = \{2\}\}$ do conjunto $\{1, 2\}$ tal que $\mathbf{x}_{I_1} = \mathbf{x} + \mathbf{h}$ e $\mathbf{x}_{I_2} = \mathbf{x}$, com $S = \{1, 2\}$, $\lambda_{I_1} = \lambda$ e $\lambda_{I_2} = 1 - \lambda$ em (1.41), tem-se, por (1.42), que $V_{Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})}(\lambda, 1 - \lambda) = \varepsilon_{Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})}(\lambda, 1 - \lambda)$, e a equação anterior pode ser reescrita na seguinte forma

$$P(Z(\mathbf{x} + \mathbf{h}) \leq z_1, Z(\mathbf{x}) \leq z_2) = F^{\varepsilon_{Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})}(\lambda, 1 - \lambda)}(z_1 + z_2).$$

Esta abordagem conduz à definição do λ -madograma de Naveau et al.(2009) [34], que caracteriza completamente a estrutura de dependência bivariada de um campo estacionário max-estável simples.

Definição 22. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório max-estável estacionário com f.d. marginais univariadas iguais a F , com F Fréchet unitária. O λ -madograma define-se por*

$$\nu_F(\lambda, \mathbf{h}) = \frac{1}{2} E \left[\left| F^\lambda(Z(\mathbf{x} + \mathbf{h})) - F^{1-\lambda}(Z(\mathbf{x})) \right| \right], \quad (3.28)$$

para $\lambda \in (0, 1)$, $\mathbf{x} \in \mathcal{D}$, $\mathbf{h} \in \mathbb{R}^d$ e $\mathbf{x} + \mathbf{h} \in \mathcal{D}$.

Na próxima proposição, estabelece-se a relação entre o λ -madograma e a função coeficiente extremal regional definida em (1.41), destacando o papel do madograma como medida de dependência.

Proposição 15. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório max-estável estacionário, com f.d. marginal univariada F Fréchet unitária. Então, a seguinte relação é válida*

$$\nu_F(\lambda, \mathbf{h}) = \frac{\varepsilon_{Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})}(\lambda, 1-\lambda)}{\varepsilon_{Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})}(\lambda, 1-\lambda) + 1} - \frac{3}{2(1+\lambda)(2-\lambda)}, \quad (3.29)$$

para $\lambda \in (0, 1)$, $\mathbf{x} \in \mathcal{D}$, $\mathbf{h} \in \mathbb{R}^d$ e $\mathbf{x} + \mathbf{h} \in \mathcal{D}$.

Demonstração. Aplicando a igualdade

$$\frac{|a-b|}{2} = \{a \vee b\} - \frac{a+b}{2}$$

à definição (3.28) do λ -madograma, obtém-se

$$\nu_F(\lambda, \mathbf{h}) = E \left[F^\lambda(Z(\mathbf{x} + \mathbf{h})) \vee F^{1-\lambda}(Z(\mathbf{x})) \right] - \frac{1}{2} \left(E \left[F^\lambda(Z(\mathbf{x} + \mathbf{h})) \right] + E \left[F^{1-\lambda}(Z(\mathbf{x})) \right] \right). \quad (3.30)$$

Considerando a partição $\mathcal{I} = \{I_1 = \{1\}, I_2 = \{2\}\}$ do conjunto $\{1, 2\}$, $\lambda_{I_1} = \lambda$, $\lambda_{I_2} = 1 - \lambda$, com $\lambda \in (0, 1)$, $S = \{1, 2\}$, $\mathbf{x}_1 = \mathbf{x} + \mathbf{h}$ e $\mathbf{x}_2 = \mathbf{x}$, por (1.43), tem-se

$$E \left[F^\lambda(Z(\mathbf{x} + \mathbf{h})) \vee F^{1-\lambda}(Z(\mathbf{x})) \right] = \frac{\varepsilon_{Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})}(\lambda, 1-\lambda)}{1 + \varepsilon_{Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})}(\lambda, 1-\lambda)}. \quad (3.31)$$

Considerando agora a partição $\mathcal{I} = \{I_1 = \{1\}\}$ do conjunto $\{1\}$, $\lambda_{I_1} = \lambda$, com $\lambda \in (0, 1)$, $S = \{1\}$ e $\mathbf{x}_1 = \mathbf{x} + \mathbf{h}$, segue-se, por (1.44), que

$$E \left[F^\lambda(Z(\mathbf{x} + \mathbf{h})) \right] = \frac{1}{\lambda + 1}, \quad (3.32)$$

uma vez que $\varepsilon_{Z(\mathbf{x}+\mathbf{h})}(1) = 1$.

Usando argumentos similares, deduz-se que

$$E \left[F^{1-\lambda}(Z(\mathbf{x})) \right] = \frac{1}{2-\lambda}. \quad (3.33)$$

Substituindo (3.31), (3.32) e (3.33) em (3.30), obtém-se

$$\begin{aligned} \nu_F(\lambda, \mathbf{h}) &= \frac{\varepsilon_{Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})}(\lambda, 1-\lambda)}{\varepsilon_{Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})}(\lambda, 1-\lambda) + 1} - \frac{1}{2} \left(\frac{1}{\lambda + 1} + \frac{1}{2-\lambda} \right) \\ &= \frac{\varepsilon_{Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})}(\lambda, 1-\lambda)}{\varepsilon_{Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})}(\lambda, 1-\lambda) + 1} - \frac{3}{2(1+\lambda)(2-\lambda)}. \end{aligned}$$

□

Com base na proposição anterior, obtém-se o enquadramento do λ -madograma dado na próxima proposição.

Proposição 16. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório max-estável estacionário com f.d. marginal univariada F Fréchet unitária. Então, para $\lambda \in (0, 1)$, tem-se*

$$\left| \frac{1 - 2\lambda}{2(1 + \lambda)(2 - \lambda)} \right| \leq \nu_F(\lambda, \mathbf{h}) \leq \frac{1}{\lambda(1 - \lambda) + 1} - \frac{3}{2(1 + \lambda)(2 - \lambda)}, \quad (3.34)$$

sendo que os limites inferior e superior correspondem, respectivamente, aos valores que o madograma assume sob as condições de dependência total e independência entre as margens $Z(\mathbf{x} + \mathbf{h})$ e $Z(\mathbf{x})$, para $\lambda \in (0, 1)$, $\mathbf{x} \in \mathcal{D}$ e $\mathbf{h} \in \mathbb{R}^d$ tais que $\mathbf{x} + \mathbf{h} \in \mathcal{D}$.

Demonstração.

Pela expressão (3.29) e tendo em conta (1.46), se se considerar a partição $\mathcal{I} = \{I_1 = \{1\}, I_2 = \{2\}\}$ do conjunto $\{1, 2\}$, $\lambda_{I_1} = \lambda$, $\lambda_{I_2} = 1 - \lambda$, com $\lambda \in (0, 1)$, $S = \{1, 2\}$, $\mathbf{x}_1 = \mathbf{x} + \mathbf{h}$ e $\mathbf{x}_2 = \mathbf{x}$, deriva-se o seguinte intervalo de variação para $\nu_F(\lambda, \mathbf{h})$:

$$\frac{\frac{1}{\lambda \wedge (1 - \lambda)}}{1 + \frac{1}{\lambda \wedge (1 - \lambda)}} - \frac{3}{2(1 + \lambda)(2 - \lambda)} \leq \nu_F(\lambda, \mathbf{h}) \leq \frac{\frac{1}{\lambda(1 - \lambda)}}{1 + \frac{1}{\lambda(1 - \lambda)}} - \frac{3}{2(1 + \lambda)(2 - \lambda)},$$

onde os limites inferior e superior correspondem, respectivamente, aos casos de dependência total e independência entre $Z(\mathbf{x} + \mathbf{h})$ e $Z(\mathbf{x})$. Mas,

$$\begin{aligned} \frac{\frac{1}{\lambda \wedge (1 - \lambda)}}{1 + \frac{1}{\lambda \wedge (1 - \lambda)}} - \frac{3}{2(1 + \lambda)(2 - \lambda)} &= \begin{cases} \frac{1/\lambda}{1/\lambda + 1} - \frac{3}{2(1 + \lambda)(2 - \lambda)}, & \text{se } \lambda \wedge (1 - \lambda) = \lambda, \\ \frac{1/(1 - \lambda)}{1/(1 - \lambda) + 1} - \frac{3}{2(1 + \lambda)(2 - \lambda)}, & \text{se } \lambda \wedge (1 - \lambda) = 1 - \lambda. \end{cases} \\ &= \begin{cases} \frac{1 - 2\lambda}{2(1 + \lambda)(2 - \lambda)}, & \text{se } \lambda < 1/2, \\ \frac{-1 + 2\lambda}{2(1 + \lambda)(2 - \lambda)}, & \text{se } \lambda > 1/2. \end{cases} \\ &= \left| \frac{1 - 2\lambda}{2(1 + \lambda)(2 - \lambda)} \right|; \end{aligned}$$

e

$$\frac{\frac{1}{\lambda(1 - \lambda)}}{\frac{1}{\lambda(1 - \lambda)} + 1} - \frac{3}{2(1 + \lambda)(2 - \lambda)} = \frac{1}{\lambda(1 - \lambda) + 1} - \frac{3}{2(1 + \lambda)(2 - \lambda)},$$

logo,

$$\left| \frac{1 - 2\lambda}{2(1 + \lambda)(2 - \lambda)} \right| \leq \nu_F(\lambda, \mathbf{h}) \leq \frac{1}{\lambda(1 - \lambda) + 1} - \frac{3}{2(1 + \lambda)(2 - \lambda)},$$

correspondendo os limites inferior e superior, respectivamente, aos casos de dependência total e independência entre as margens $Z(\mathbf{x} + \mathbf{h})$ e $Z(\mathbf{x})$. □

As constantes λ e $1 - \lambda$ representam as ponderações atribuídas às variáveis $F(Z(\mathbf{x} + \mathbf{h}))$ e $F(Z(\mathbf{x}))$, respetivamente. Intuitivamente, podemos pensar nelas como as ponderações atribuídas às localizações $\mathbf{x} + \mathbf{h}$ e \mathbf{x} .

Exemplo 12. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ um campo aleatório max-estável estacionário com f.d. marginal univariada F Fréchet unitária.*

Ao fixar $\lambda = 1/2$ em $\nu_F(\lambda, \mathbf{h})$, atribuem-se pesos iguais a $1/2$ a $F(Z(\mathbf{x} + \mathbf{h}))$ e a $F(Z(\mathbf{x}))$, resultando numa versão do λ -madograma que se assemelha à do F -madograma.

Por (3.29), o $1/2$ -madograma é dado por

$$\begin{aligned} \nu_F(1/2, \mathbf{h}) &= \frac{\varepsilon_{Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})}(1/2, 1/2)}{\varepsilon_{Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})}(1/2, 1/2) + 1} - \frac{3}{2(1 + 1/2)(2 - 1/2)} \\ &= \frac{2\varepsilon_{Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})}(1, 1)}{2\varepsilon_{Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})}(1, 1) + 1} - \frac{2}{3}. \end{aligned}$$

Estimação do λ -madograma

Seja $(Z^{(t)}(\mathbf{x} + \mathbf{h}), Z^{(t)}(\mathbf{x}))$, $t \in \{1, \dots, T\}$, uma amostra aleatória de $(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x}))$, com f.d. marginais iguais a $F_{Z(\mathbf{x}+\mathbf{h})}$ e $F_{Z(\mathbf{x})}$, respetivamente. Repetindo o processo de transformação das margens $Z^{(t)}(\mathbf{x} + \mathbf{h})$ e $Z^{(t)}(\mathbf{x})$, para $t \in \{1, \dots, T\}$, em (3.24), de forma a obter Fréchet unitárias, segue-se que

$$\hat{\nu}_F(\lambda, \mathbf{h}) = \frac{1}{2T} \sum_{t=1}^T \left| F_{Z(\mathbf{x}+\mathbf{h})}^\lambda \left(Z^{(t)}(\mathbf{x} + \mathbf{h}) \right) - F_{Z(\mathbf{x})}^{1-\lambda} \left(Z^{(t)}(\mathbf{x}) \right) \right| \quad (3.35)$$

é um estimador fortemente consistente e assintoticamente normal para $\nu_F(\lambda, \mathbf{h})$ (Naveau et al. (2009) [34]). Se as f.d. marginais $F_{Z(\mathbf{x}+\mathbf{h})}$ e $F_{Z(\mathbf{x})}$ forem desconhecidas, podem ser substituídas em (3.35) pelas correspondentes f.d. empíricas corrigidas.

O λ -madograma verifica as condições de fronteira $\lim_{\lambda \rightarrow 0^+} \nu_F(\lambda, \mathbf{h}) = \frac{1}{4} = \lim_{\lambda \rightarrow 1^-} \nu_F(\lambda, \mathbf{h})$. Para que se tenha $\lim_{\lambda \rightarrow 0^+} E[\hat{\nu}_F(\lambda, \mathbf{h})] = \frac{1}{4} = \lim_{\lambda \rightarrow 1^-} E[\hat{\nu}_F(\lambda, \mathbf{h})]$, Naveau et al. (2009) [34] propuseram uma versão corrigida do estimador, dada por:

$$\begin{aligned} \hat{\hat{\nu}}_F(\lambda, \mathbf{h}) &= \hat{\nu}_F(\lambda, \mathbf{h}) - \frac{\lambda}{2T} \sum_{t=1}^T \left\{ 1 - F_{Z(\mathbf{x}+\mathbf{h})}^\lambda(Z^{(t)}(\mathbf{x} + \mathbf{h})) \right\} \\ &\quad - \frac{1-\lambda}{2T} \sum_{t=1}^T \left\{ 1 - F_{Z(\mathbf{x})}^{1-\lambda}(Z^{(t)}(\mathbf{x})) \right\} + \frac{1-\lambda+\lambda^2}{2(2-\lambda)(1+\lambda)}. \end{aligned} \quad (3.36)$$

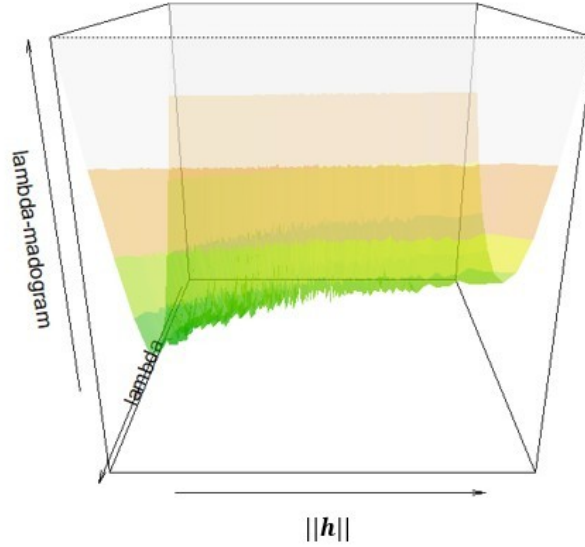


Figura 3.8: Estimativas do λ -madograma para o conjunto de dados *rainfall*.

Os termos adicionais asseguram que as condições de fronteira em $\lambda = 0$ e $\lambda = 1$ sejam cumpridas por este último estimador, tornando-o mais robusto e consistente em aplicações práticas, quando comparado com o primeiro. Aplicando-o na relação (3.29), deduz-se que

$$\hat{\varepsilon}_{Z(\mathbf{x}+\mathbf{h}),Z(\mathbf{x})}(\lambda, 1 - \lambda) = \frac{\hat{\nu}_F(\lambda, \mathbf{h}) + \frac{3}{2(1+\lambda)(2-\lambda)}}{-\hat{\nu}_F(\lambda, \mathbf{h}) - \frac{3}{2(1+\lambda)(2-\lambda)} + 1} \quad (3.37)$$

é um estimador para $\varepsilon_{Z(\mathbf{x}+\mathbf{h}),Z(\mathbf{x})}(\lambda, 1 - \lambda)$.

Exemplo 13. Aplicando agora o comando **lmadogram** da biblioteca *SpatialExtremes* do software R ao conjunto de dados apresentado no Exemplo 10, obtêm-se as estimativas para o λ -madograma apresentadas no gráfico da Figura 3.8.

Constatam-se valores crescentes do λ -madograma em função de $\|\mathbf{h}\|$, o que evidencia uma diminuição da dependência extremal à medida que aumenta a distância entre as localizações. Adicionalmente, verifica-se que o λ -madograma é sensível à distribuição dos pesos entre as va. $F(Z(x+h))$ e $F(Z(x))$.

λ -madograma multivariado

Na sequência da apresentação de madogramas que permitem mensurar a dependência bivariada em campos max-estáveis, passamos agora a explorar madogramas que capturam a dependência k -variada, onde $k \geq 2$. O primeiro que analisamos é uma extensão natural do λ -madograma, conforme será observado. De seguida, apresentamos a sua definição formal.

Definição 23. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo max-estável simples com funções de distribuição marginais univariadas iguais a F . O $(\lambda_1, \dots, \lambda_k)$ -madograma define-se por*

$$\nu_F(\lambda_1, \dots, \lambda_k) = E \left[\bigvee_{i=1}^k F^{\lambda_i}(Z(\mathbf{x}_i)) - \frac{1}{k} \sum_{i=1}^k F^{\lambda_i}(Z(\mathbf{x}_i)) \right], \quad (3.38)$$

para $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{D}$ e $\lambda_1, \dots, \lambda_k \in (0, +\infty)$.

Observação 1. *Para $k = 2$, considerando $\lambda_1 = \lambda$ e $\lambda_2 = 1 - \lambda$, com $\lambda \in (0, 1)$, e supondo estacionariedade, obtém-se o λ -madograma de Naveau et al. (2009) [34], definido em (3.28). Outro caso particular deste madograma é o de Marcon et al. (2017) [31] com distribuições marginais univariadas iguais a F , tomando $\lambda_i = \frac{1}{w_i}$, com $w_i \in [0, 1]$, para $i \in \{1, \dots, k\}$, e tais que $w_1 + \dots + w_k = 1$.*

Na próxima proposição, estabelece-se a relação entre o madograma e a função coeficiente extremal regional.

Proposição 17. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo max-estável com funções de distribuição marginais univariadas iguais a F , com F Fréchet unitária. Então*

$$\nu_F(\lambda_1, \dots, \lambda_k) = \frac{\varepsilon_{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)}(\lambda_1, \dots, \lambda_k)}{\varepsilon_{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)}(\lambda_1, \dots, \lambda_k) + 1} - \frac{1}{k} \sum_{i=1}^k \frac{1}{1 + \lambda_i}, \quad (3.39)$$

para $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{D}$ e $\lambda_1, \dots, \lambda_k \in (0, +\infty)$.

Demonstração. Aplicando as propriedades da esperança matemática em (3.38), tem-se

$$\nu_F(\lambda_1, \dots, \lambda_k) = E \left[\bigvee_{i=1}^k F^{\lambda_i}(Z(\mathbf{x}_i)) \right] - \frac{1}{k} \sum_{i=1}^k E \left[F^{\lambda_i}(Z(\mathbf{x}_i)) \right]. \quad (3.40)$$

Tendo em conta (1.43), obtém-se

$$E \left[\bigvee_{i=1}^k F^{\lambda_i}(Z(\mathbf{x}_i)) \right] = \frac{\varepsilon_{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)}(\lambda_1, \dots, \lambda_k)}{1 + \varepsilon_{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)}(\lambda_1, \dots, \lambda_k)}, \quad (3.41)$$

tomando a partição $\mathcal{I} = \{I_1 = \{1\}, \dots, I_k = \{k\}\}$ do conjunto $\{1, \dots, k\}$, $\lambda_{I_1} = \lambda_1, \dots, \lambda_{I_k} = \lambda_k$, com $\lambda_i \in (0, +\infty)$, $i \in \{1, \dots, k\}$, e $S = \{1, \dots, k\}$.

Considerando agora a partição $\mathcal{I} = \{I_i = \{i\}\}$ do conjunto $\{i\}$, $\lambda_{I_i} = \lambda_i$, com $\lambda_i \in (0, +\infty)$, e $S = \{i\}$, deriva-se, de (1.44),

$$E \left[F^{1/\lambda_i}(Z(\mathbf{x}_i)) \right] = \frac{1}{1 + \lambda_i}. \quad (3.42)$$

Substituindo (3.41) e (3.42) em (3.40), deduz-se o pretendido:

$$\nu_F(\lambda_1, \dots, \lambda_k) = \frac{\varepsilon_{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)}(\lambda_1, \dots, \lambda_k)}{\varepsilon_{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)}(\lambda_1, \dots, \lambda_k) + 1} - \frac{1}{k} \sum_{i=1}^k \frac{1}{1 + \lambda_i}.$$

□

Aplicamos a proposição anterior para calcular o madograma de um vetor proveniente de um campo de máximos multivariados de máximos móveis (do inglês *'Multivariate Maxima of Moving Maxima'*). Estes campos, também designados por campos $M4$, foram introduzidos por Smith e Weissman (1996) [45] como uma classe de campos construídos a partir de máximos de variáveis independentes ponderados por coeficientes não negativos.

Exemplo 14. Considere-se $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^2}$ um campo aleatório $M4$ definido por

$$Z(\mathbf{x}) = \bigvee_{l \in \mathbb{N}} \bigvee_{m \in \mathbb{Z}} a_{l,m,\mathbf{x}} Y_{l,1-m}, \quad (3.43)$$

com $\{Y_{l,1-m}\}_{l \in \mathbb{N}, m \in \mathbb{Z}}$ uma coleção de variáveis aleatórias independentes com funções de distribuição marginais univariadas iguais a F , com F Fréchet unitária, e com $\{a_{l,m,\mathbf{x}}\}_{l \in \mathbb{N}, m \in \mathbb{Z}, \mathbf{x} \in \mathbb{R}^2}$ uma coleção de constantes não negativas satisfazendo

$$\sum_{l \in \mathbb{N}} \sum_{m \in \mathbb{Z}} a_{l,m,\mathbf{x}} = 1.$$

Estas constantes designam-se padrões de assinaturas, sendo l o número de padrões e m o número de assinaturas.

Em Fonseca et al. (2016) [20], prova-se que o campo $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ é max-estável.

A função de distribuição do vetor $\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$ é igual a

$$\begin{aligned} & P(Z(\mathbf{x}_1) \leq z_1, \dots, Z(\mathbf{x}_k) \leq z_k) \\ &= P \left(\bigvee_{l \in \mathbb{N}} \bigvee_{m \in \mathbb{Z}} a_{l,m,\mathbf{x}_1} Y_{l,1-m} \leq z_1, \dots, \bigvee_{l \in \mathbb{N}} \bigvee_{m \in \mathbb{Z}} a_{l,m,\mathbf{x}_k} Y_{l,1-m} \leq z_k \right) \\ &= P \left(\bigcap_{l \in \mathbb{N}} \left\{ \bigcap_{m \in \mathbb{Z}} \{a_{l,m,\mathbf{x}_1} Y_{l,1-m} \leq z_1\} \right\}, \dots, \bigcap_{l \in \mathbb{N}} \left\{ \bigcap_{m \in \mathbb{Z}} \{a_{l,m,\mathbf{x}_k} Y_{l,1-m} \leq z_k\} \right\} \right) \\ &= P \left(\bigcap_{l \in \mathbb{N}} \left\{ \bigcap_{m \in \mathbb{Z}} \left\{ Y_{l,1-m} \leq \bigwedge_{i=1}^k \frac{z_i}{a_{l,m,\mathbf{x}_i}} \right\} \right\} \right) \end{aligned}$$

$$\begin{aligned}
&= \prod_{l \in \mathbb{N}} \prod_{m \in \mathbb{Z}} P \left(Y_{l,1-m} \leq \bigwedge_{i=1}^k \frac{z_i}{a_{l,m,\mathbf{x}_i}} \right) \\
&= \prod_{l \in \mathbb{N}} \prod_{m \in \mathbb{Z}} \exp \left(- \frac{1}{\bigwedge_{i=1}^k \frac{z_i}{a_{l,m,\mathbf{x}_i}}} \right) \\
&= \prod_{l \in \mathbb{N}} \prod_{m \in \mathbb{Z}} \exp \left(- \bigvee_{i=1}^k \frac{a_{l,m,\mathbf{x}_i}}{z_i} \right) \\
&= \exp \left(- \sum_{l \in \mathbb{N}} \sum_{m \in \mathbb{Z}} \bigvee_{i=1}^k \frac{a_{l,m,\mathbf{x}_i}}{z_i} \right), \quad z_1, \dots, z_k \in \mathbb{R}^+,
\end{aligned}$$

uma vez que as variáveis $Y_{l,1-m}$ são independentes e têm distribuição comum Fréchet unitária.

Por (3.5),

$$P(Z(\mathbf{x}_1) \leq z_1, \dots, Z(\mathbf{x}_k) \leq z_k) = \exp(-V_{\mathbf{Z}}(z_1, \dots, z_k)).$$

Como, por (1.42), $V_{\mathbf{Z}}(z_1, \dots, z_k) = \varepsilon_{\mathbf{Z}}(z_1, \dots, z_k)$, segue-se que

$$\varepsilon_{\mathbf{Z}}(z_1, \dots, z_k) = \sum_{l \in \mathbb{N}} \sum_{m \in \mathbb{Z}} \bigvee_{i=1}^k \frac{a_{l,m,\mathbf{x}_i}}{z_i}. \quad (3.44)$$

Considerando, em particular, para $\mathbf{x}_i \in \mathbb{Z}^2$, um padrão, $l = 1$, de duas assinaturas, $1 \leq m \leq 2$, com valores

$$\left\{ \begin{array}{ll} a_{1,1,\mathbf{x}_i} = a_{1,2,\mathbf{x}_i} = \frac{1}{2}, & \text{se ambas as coordenadas de } \mathbf{x}_i \text{ são pares;} \\ 1 - a_{1,1,\mathbf{x}_i} = a_{1,2,\mathbf{x}_i} = \frac{1}{3}, & \text{se ambas as coordenadas de } \mathbf{x}_i \text{ são ímpares;} \\ a_{1,1,\mathbf{x}_i} = \frac{1}{3} = 1 - a_{1,2,\mathbf{x}_i}, & \text{se uma coordenada de } \mathbf{x}_i \text{ é par e outra é ímpar;} \end{array} \right.$$

apliquemos a Proposição 17 para calcular o valor do madograma para $\mathbf{Z} = (Z(\mathbf{x}_1), Z(\mathbf{x}_2), Z(\mathbf{x}_3))$, com $\mathbf{x}_1 = (1, 1)$, $\mathbf{x}_2 = (2, 1)$ e $\mathbf{x}_3 = (2, 2)$.

Na Figura 3.9, encontra-se uma simulação do campo M4.

De (3.44), deriva-se que

$$\begin{aligned}
\varepsilon_{Z(\mathbf{x}_1), Z(\mathbf{x}_2), Z(\mathbf{x}_3)}(\lambda_1, \lambda_2, \lambda_3) &= \sum_{m=1}^2 \bigvee_{i=1}^3 \frac{a_{1,m,\mathbf{x}_i}}{\lambda_i} \\
&= \bigvee_{i=1}^3 \frac{a_{1,1,\mathbf{x}_i}}{\lambda_i} + \bigvee_{i=1}^3 \frac{a_{1,2,\mathbf{x}_i}}{\lambda_i}
\end{aligned}$$

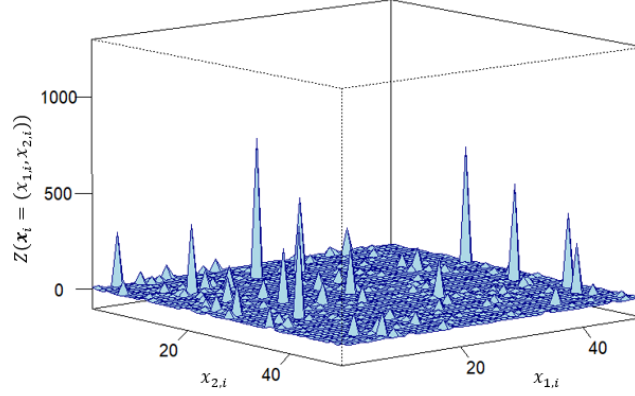


Figura 3.9: Simulação do campo $M4$ apresentado no Exemplo 14.

$$\begin{aligned}
&= \frac{a_{1,1,(1,1)}}{\lambda_1} \vee \frac{a_{1,1,(2,1)}}{\lambda_2} \vee \frac{a_{1,1,(2,2)}}{\lambda_3} \\
&+ \frac{a_{1,2,(1,1)}}{\lambda_1} \vee \frac{a_{1,2,(2,1)}}{\lambda_2} \vee \frac{a_{1,2,(2,2)}}{\lambda_3} \\
&= \frac{2}{3\lambda_1} \vee \frac{1}{3\lambda_2} \vee \frac{1}{2\lambda_3} + \frac{1}{3\lambda_1} \vee \frac{2}{3\lambda_2} \vee \frac{1}{2\lambda_3},
\end{aligned}$$

para $\lambda_1, \lambda_2, \lambda_3 \in (0, +\infty)$. Substituindo em (3.39), conclui-se que

$$\nu_F(\lambda_1, \lambda_2, \lambda_3) = \frac{\frac{2}{3\lambda_1} \vee \frac{1}{3\lambda_2} \vee \frac{1}{2\lambda_3} + \frac{1}{3\lambda_1} \vee \frac{2}{3\lambda_2} \vee \frac{1}{2\lambda_3}}{\frac{2}{3\lambda_1} \vee \frac{1}{3\lambda_2} \vee \frac{1}{2\lambda_3} + \frac{1}{3\lambda_1} \vee \frac{2}{3\lambda_2} \vee \frac{1}{2\lambda_3} + 1} - \frac{1}{3} \sum_{i=1}^3 \frac{1}{1 + \lambda_i}.$$

A partir da relação (3.39) do madograma com a função coeficiente extremal regional da Proposição 17, é possível derivar o intervalo de variação do madograma estabelecido na próxima proposição.

Proposição 18. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo max-estável com funções de distribuição marginais univariadas, iguais a F , Fréchet unitárias. Então*

$$\frac{\bigvee_{i=1}^k \frac{1}{\lambda_i}}{\bigvee_{i=1}^k \frac{1}{\lambda_i} + 1} - \frac{1}{k} \sum_{i=1}^k \frac{1}{1 + \lambda_i} \leq \nu_F(\lambda_1, \dots, \lambda_k) \leq \frac{\sum_{i=1}^k \frac{1}{\lambda_i}}{\sum_{i=1}^k \frac{1}{\lambda_i} + 1} - \frac{1}{k} \sum_{i=1}^k \frac{1}{1 + \lambda_i}, \quad (3.45)$$

onde os limites inferior e superior são atingidos, respetivamente, quando as variáveis $Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)$ são totalmente dependentes e independentes, para $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{D}$ e $\lambda_1, \dots, \lambda_k \in (0, +\infty)$.

Demonstração. O resultado obtém-se facilmente, tendo em conta a relação (3.39) e o enquadramento (1.46), assumindo-se a partição $\mathcal{I} = \{I_1 = \{1\}, \dots, I_k = \{k\}\}$ de $\{1, \dots, k\}$, $S = \{1, \dots, k\}$ e $\lambda_{I_j} = \lambda_j$, com $j = 1, \dots, k$. □

Estimação do λ -madograma multivariado

Seja $(Z^{(t)}(\mathbf{x}_1), \dots, Z^{(t)}(\mathbf{x}_k))$, $t \in \{1, \dots, T\}$, uma amostra aleatória de $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$, com $F_{Z(\mathbf{x}_i)}$ a f.d. de $Z^{(t)}(\mathbf{x}_i)$, onde $t \in \{1, \dots, T\}$ e $i \in \{1, \dots, k\}$. Repetindo o processo de transformação das margens em (3.24), agora para $Z^{(t)}(\mathbf{x}_i)$, com $i \in \{1, \dots, k\}$ e $t \in \{1, \dots, T\}$, de forma a obter distribuições marginais univariadas Fréchet unitárias, segue-se que

$$\hat{\nu}_F(\lambda_1, \dots, \lambda_k) = \frac{1}{T} \sum_{t=1}^T \left\{ \prod_{i=1}^k F_{Z(\mathbf{x}_i)}^{\lambda_i}(Z^{(t)}(\mathbf{x}_i)) - \frac{1}{k} \sum_{i=1}^k F_{Z(\mathbf{x}_i)}^{\lambda_i}(Z^{(t)}(\mathbf{x}_i)) \right\}, \quad (3.46)$$

é um estimador para $\nu_F(\lambda_1, \dots, \lambda_k)$, com $\lambda_1, \dots, \lambda_k \in (0, +\infty)$. Quando as distribuições $F_{Z(\mathbf{x}_i)}$, $i \in \{1, \dots, k\}$, são desconhecidas, substituem-se pelas correspondentes empíricas corrigidas. Da relação (3.39), obtém-se que

$$\hat{\varepsilon}_{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)}(\lambda_1, \dots, \lambda_k) = \frac{\hat{\nu}_F(\lambda_1, \dots, \lambda_k) + \frac{1}{k} \sum_{i=1}^k \frac{1}{1+\lambda_i}}{-\hat{\nu}_F(\lambda_1, \dots, \lambda_k) - \frac{1}{k} \sum_{i=1}^k \frac{1}{1+\lambda_i} + 1} \quad (3.47)$$

é um estimador para $\varepsilon_{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)}(\lambda_1, \dots, \lambda_k)$.

Exemplo 15. *Com o objetivo de investigar se a dependência entre precipitações extremas do conjunto de dados **rainfall**, apresentado no Exemplo 10, é mais acentuada em localizações geograficamente mais homogêneas, aplicámos a implementação dos estimadores anteriores no R a dois subconjuntos de estações meteorológicas:*

- *Conjunto A: localizações l_5, l_{13}, l_{15} e l_{19} , situadas numa região relativamente plana, com altitudes semelhantes e características orográficas semelhantes;*
- *Conjunto B: localizações l_{43}, l_{53}, l_{57} e l_{75} , situadas em zonas montanhosas, caracterizadas por maior variabilidade altitudinal e potencial influência de microclimas diversificados.*

As latitudes, longitudes e altitudes das estações selecionadas são apresentadas na Tabela 3.1.

Não existindo razões para privilegiar nenhuma localização, assumiram-se ponderações iguais para todas. Ainda que, teoricamente, fosse possível atribuir diferentes ponderações às localizações — explorando, por exemplo, cenários em que certas estações contribuem mais significativamente para a medida de dependência — tal abordagem implicaria considerar um número elevado de combinações, o que poderia tornar a análise excessivamente

Tabela 3.1: Longitudes, latitudes e altitudes do subconjunto em estudo no Exemplo 15.

Localização	Longitude (m)	Latitude (m)	Altitude (m)
l_5	693.810	273.070	382
l_{13}	682.010	263.780	402
l_{15}	696.730	283.130	405
l_{19}	695.350	255.060	480
l_{43}	730.875	219.535	687
l_{53}	744.200	234.920	2502
l_{57}	710.800	225.520	452
l_{75}	746.130	229.570	1000

complexa e de interpretação menos direta. A Figura 3.10 apresenta estimativas do madograma para $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda \in (0, 100)$, juntamente com os limites teóricos inferior e superior.

Para o caso particular $\lambda_1 = 1$, calcularam-se as estimativas do madograma e da função coeficiente extremal, apresentadas na Tabela 3.2. De acordo com os enquadramentos teóricos dados em (3.45) e (1.46), tem-se

$$0 \leq \nu_F(1, 1, 1, 1) \leq 1/3$$

e

$$1 \leq \varepsilon_{Z(\mathbf{x}_1), Z(\mathbf{x}_2), Z(\mathbf{x}_3), Z(\mathbf{x}_4)}(1, 1, 1, 1) \leq 5,$$

onde os limites inferiores e superiores correspondem, respectivamente, aos casos de dependência total e independência.

Tabela 3.2: Estimativas do madograma $\nu_F(1, 1, 1, 1)$ e da função coeficiente extremal $\varepsilon_{Z(\mathbf{x}_1), Z(\mathbf{x}_2), Z(\mathbf{x}_3), Z(\mathbf{x}_4)}(1, 1, 1, 1)$ para os conjuntos A e B do Exemplo 15.

Conjunto	$\hat{\nu}_F(1, 1, 1, 1)$	$\hat{\varepsilon}_{Z(\mathbf{x}_1), Z(\mathbf{x}_2), Z(\mathbf{x}_3), Z(\mathbf{x}_4)}(1, 1, 1, 1)$
A	0.1575	1.9194
B	0.2200	2.5710

Observa-se, então, que as estimativas associadas ao Conjunto A, tanto para o madograma como para a função coeficiente extremal, são inferiores às obtidas para o Conjunto B, aproximando-se mais dos respectivos limites inferiores teóricos. Isto sugere uma dependência mais forte entre as precipitações extremas nas localizações do Conjunto A, em comparação com as do Conjunto B, o que está de acordo com o expectável, dada a maior homogeneidade orográfica e altitudinal do Conjunto A. Note-se ainda que, para valores muito elevados de λ , as estimativas do madograma para ambos os conjuntos convergem para valores muito próximos de zero.

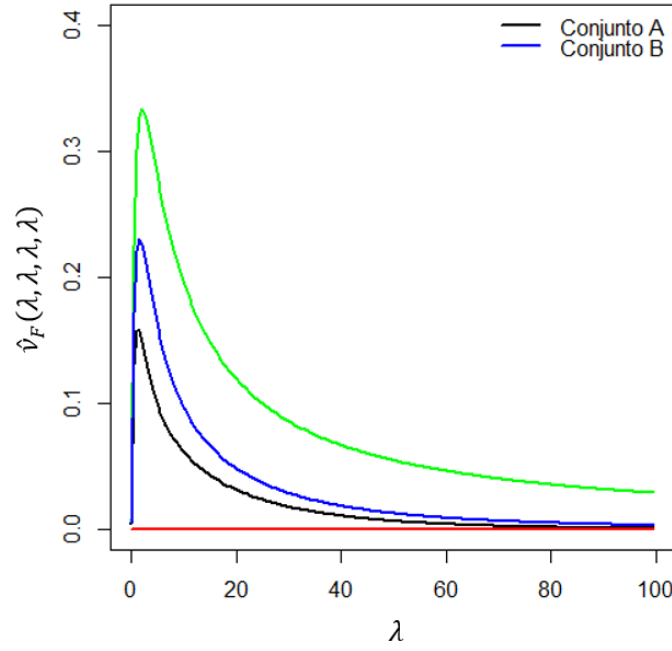


Figura 3.10: Estimativas de $\nu_F(\lambda, \lambda, \lambda, \lambda)$, com $\lambda \in (0, 100)$, para o Conjunto A (a preto) e o Conjunto B (a azul), considerados no Exemplo 15, e limites teóricos inferior (a vermelho, igual a zero) e superior (a verde, igual a $1/(1 + \lambda/4) - 1/(1 + \lambda)$).

Madograma generalizado

O madograma sobre o qual nos debruçaremos de seguida — madograma generalizado — foi introduzido por Fonseca (2012) [19] e permite, também, mensurar a dependência multivariada de um campo max-estável, mas para variáveis indexadas em duas regiões de localizações disjuntas. Mais especificamente, este madograma possibilita a análise da dependência entre os máximos das variáveis indexadas nessas duas regiões. A relevância desta análise advém do facto de os hidrologistas agruparem os dados em regiões com base nas suas características geográficas e hidrográficas, conforme mencionado por Fonseca (2012) [19].

Definição 24. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório max-estável simples com funções de distribuição marginais univariadas iguais a F . Considerando dois conjuntos disjuntos $I_1 = \{i_0 + 1, \dots, i_1\}$ e $I_2 = \{i_1 + 1, \dots, i_2\}$, com $0 = i_0 < i_1 < i_2$ e $i_1, i_2 \in \mathbb{N}$, tais que $\mathbf{x}_i \in \mathcal{D}$, para $i \in \{1, \dots, i_2\}$, o madograma generalizado é dado por*

$$\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2)) := \frac{1}{2} E \left[\left| F^{\lambda_1}(M(\mathbf{Z}_{I_1})) - F^{\lambda_2}(M(\mathbf{Z}_{I_2})) \right| \right], \quad (3.48)$$

para $\lambda_1, \lambda_2 \in \mathbb{R}^+$, onde $\mathbf{Z} = (\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2})$, com $\mathbf{Z}_{I_1} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_{i_1}))$, $\mathbf{Z}_{I_2} = (Z(\mathbf{x}_{i_1+1}), \dots, Z(\mathbf{x}_{i_2}))$, $M(\mathbf{Z}_{I_1}) = \bigvee_{i \in I_1} Z(\mathbf{x}_i) = \bigvee_{i=1}^{i_1} Z(\mathbf{x}_i)$ e $M(\mathbf{Z}_{I_2}) = \bigvee_{i \in I_2} Z(\mathbf{x}_i) = \bigvee_{i=i_1+1}^{i_2} Z(\mathbf{x}_i)$.

Observação 2. As constantes λ_1 e λ_2 , que surgem na definição anterior, representam os pesos atribuídos às variáveis $F(M(\mathbf{Z}_{I_1}))$ e $F(M(\mathbf{Z}_{I_2}))$, ou, com um certo abuso de linguagem, as ponderações atribuídas às regiões de localizações indexadas em I_1 e I_2 . Relevamos três escolhas possíveis:

(i) Para $\lambda_1 = \lambda_2 = 1$, $I_1 = \{1\}$ e $I_2 = \{2\}$ tais que $x_1 = x + h$ e $x_2 = x$, o madograma generalizado reduz-se ao F -madograma (3.19), supondo que o campo $\{Z(x)\}_{x \in \mathcal{D}}$ é estacionário;

(ii) Quando se considera $\lambda_2 = 1 - \lambda_1$, $\lambda_1 \in (0, 1)$, $I_1 = \{1\}$ e $I_2 = \{2\}$, tais que $x_1 = x + h$ e $x_2 = x$ em (3.48) e se supõe estacionariedade, obtém-se a expressão (3.28). Assim, o madograma generalizado reduz-se ao λ -madograma, com $\lambda = \lambda_1$;

(iii) Para estender a abordagem do λ -madograma e assegurando que nenhuma das regiões de localizações indexadas em I_1 e I_2 é privilegiada em função do seu número de localizações, pode-se considerar

$$\lambda_1 = \frac{\lambda}{i_1} \text{ e } \lambda_2 = \frac{1 - \lambda}{i_2 - i_1}, \text{ com } \lambda \in (0, 1).$$

Tendo em conta que F é Fréchet unitária, i.e., $F(z) = \exp(-\frac{1}{z})$, para $z > 0$, a expressão (3.48) pode ser manipulada de forma a obter

$$\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2)) = \frac{1}{2} E \left[\left| \prod_{i \in I_1} F^{\lambda_1}(Z(x_i)) - \prod_{i \in I_2} F^{\lambda_2}(Z(x_i)) \right| \right], \quad (3.49)$$

já que F é monótona crescente e, por isso, comuta com o máximo. Esta identidade motivou a definição do estimador proposto por Fonseca (2012) [19], a apresentar mais à frente, e revela-se essencial para a demonstração da proposição seguinte, onde se encontra uma expressão do madograma generalizado à custa da função coeficiente extremal regional subjacente ao campo aleatório. Tal formulação viabiliza o uso do madograma como ferramenta para avaliar a dependência entre os vetores indexados nas regiões de localizações com índices em I_1 e I_2 .

Proposição 19. Seja $\{Z(x)\}_{x \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório max-estável com funções de distribuição marginais univariadas iguais a F , com F Fréchet unitária. Então, para os conjuntos disjuntos $I_1 = \{i_0 + 1, \dots, i_1\}$ e $I_2 = \{i_1 + 1, \dots, i_2\}$, com $0 = i_0 < i_1 < i_2$ e $i_1, i_2 \in \mathbb{N}$, tais que $x_i \in \mathcal{D}$, com $i \in \{1, \dots, i_2\}$, e para $\lambda_1, \lambda_2 \in \mathbb{R}^+$, tem-se

$$\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2)) = \frac{\varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(\lambda_1, \lambda_2)}{1 + \varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(\lambda_1, \lambda_2)} - c(\lambda_1, \lambda_2), \quad (3.50)$$

com

$$c(\lambda_1, \lambda_2) = \frac{1}{2} \left(\frac{\varepsilon_{\mathbf{Z}_{I_1}}(1)}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1}}(1)} + \frac{\varepsilon_{\mathbf{Z}_{I_2}}(1)}{\lambda_2 + \varepsilon_{\mathbf{Z}_{I_2}}(1)} \right).$$

Demonstração. Aplicando a igualdade $|a - b| = 2(a \vee b) - (a + b)$ a (3.49), segue-se que

$$\begin{aligned} \nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2)) &= E \left[\bigvee_{i \in I_1} F^{\lambda_1}(Z(\mathbf{x}_i)) \vee \bigvee_{i \in I_2} F^{\lambda_2}(Z(\mathbf{x}_i)) \right] \\ &\quad - \frac{1}{2} E \left[\bigvee_{i \in I_1} F^{\lambda_1}(Z(\mathbf{x}_i)) \right] - \frac{1}{2} E \left[\bigvee_{i \in I_2} F^{\lambda_2}(Z(\mathbf{x}_i)) \right], \end{aligned} \quad (3.51)$$

para $\mathbf{x}_1, \dots, \mathbf{x}_{i_2} \in \mathcal{D}$.

Fixando a partição $\mathcal{I} = \{I_1 = \{1, \dots, i_1\}, I_2 = \{i_1 + 1, \dots, i_2\}\}$ do conjunto $\{1, \dots, i_2\}$, $\lambda_{I_1} = \lambda_1, \lambda_{I_2} = \lambda_2$ e $S = \{1, 2\}$ em (1.43), segue-se que

$$E \left[\bigvee_{i \in I_1} F^{\lambda_1}(Z(\mathbf{x}_i)) \vee \bigvee_{i \in I_2} F^{\lambda_2}(Z(\mathbf{x}_i)) \right] = \frac{\varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(\lambda_1, \lambda_2)}{1 + \varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(\lambda_1, \lambda_2)}. \quad (3.52)$$

Atendendo agora a (1.44),

$$E \left[\bigvee_{i \in I_1} F^{\lambda_1}(Z(\mathbf{x}_i)) \right] = \frac{\varepsilon_{\mathbf{Z}_{I_1}}(1)}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1}}(1)}. \quad (3.53)$$

e

$$E \left[\bigvee_{i \in I_2} F^{\lambda_2}(Z(\mathbf{x}_i)) \right] = \frac{\varepsilon_{\mathbf{Z}_{I_2}}(1)}{\lambda_2 + \varepsilon_{\mathbf{Z}_{I_2}}(1)}. \quad (3.54)$$

Aplicando as expressões (3.52), (3.53) e (3.54) em (3.51) obtém-se o pretendido. \square

No próximo exemplo, aplicamos a proposição anterior a duas regiões de um campo $M4$.

Exemplo 16. Calculemos o valor do madograma para o campo $M4$ introduzido no Exemplo 14, considerando $I_1 = \{1, \dots, 8\}$ e $I_2 = \{9, \dots, 16\}$, em que os pontos \mathbf{x}_i , com $i \in \{1, \dots, 16\}$ têm as coordenadas apresentadas na Figura 3.11. Ora, por (3.44), tem-se que

$$\begin{aligned} \varepsilon_{\mathbf{Z}_{I_1}}(1) &= \sum_{m=1}^2 \bigvee_{i=1}^8 a_{1,m,\mathbf{x}_i} \\ &= \bigvee_{i=1}^8 a_{1,1,\mathbf{x}_i} + \bigvee_{i=1}^8 a_{1,2,\mathbf{x}_i} \\ &= \left(\frac{2}{3} \vee \frac{1}{3} \vee \frac{1}{3} \vee \frac{1}{2} \vee \frac{2}{3} \vee \frac{1}{3} \vee \frac{1}{3} \vee \frac{1}{2} \right) \\ &\quad + \left(\frac{1}{3} \vee \frac{2}{3} \vee \frac{2}{3} \vee \frac{1}{2} \vee \frac{1}{3} \vee \frac{2}{3} \vee \frac{2}{3} \vee \frac{1}{2} \right) \\ &= \frac{4}{3}, \end{aligned}$$

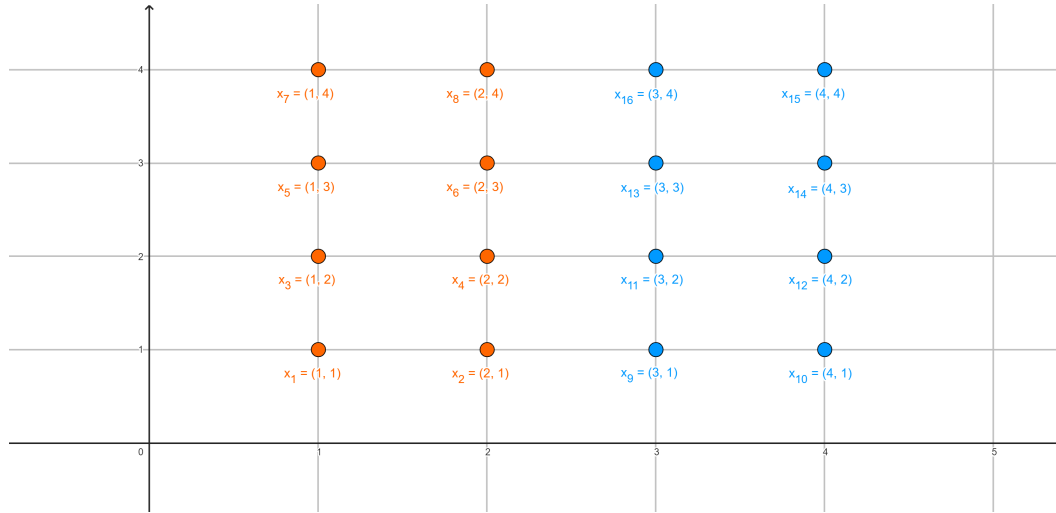


Figura 3.11: Coordenadas dos pontos $x_i, i \in \{1, \dots, 16\}$, coloridos a laranja os indexados em $I_1 = \{1, \dots, 8\}$ e a azul os em $I_2 = \{9, \dots, 16\}$.

$$\begin{aligned}
\varepsilon_{\mathbf{Z}_{I_2}}(1) &= \sum_{m=1}^2 \bigvee_{i=9}^{16} a_{1,m,x_i} \\
&= \bigvee_{i=9}^{16} a_{1,1,x_i} + \bigvee_{i=9}^{16} a_{1,2,x_i} \\
&= \left(\frac{2}{3} \vee \frac{1}{3} \vee \frac{1}{3} \vee \frac{1}{2} \vee \frac{2}{3} \vee \frac{1}{3} \vee \frac{1}{3} \vee \frac{1}{2} \right) \\
&\quad + \left(\frac{1}{3} \vee \frac{2}{3} \vee \frac{2}{3} \vee \frac{1}{2} \vee \frac{1}{3} \vee \frac{2}{3} \vee \frac{2}{3} \vee \frac{1}{2} \right) \\
&= \frac{4}{3},
\end{aligned}$$

e

$$\begin{aligned}
\varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(\lambda_1, \lambda_2) &= \sum_{m=1}^2 \left(\left(\frac{1}{\lambda_1} \bigvee_{i=1}^8 a_{1,m,x_i} \right) \vee \left(\frac{1}{\lambda_2} \bigvee_{i=9}^{16} a_{1,m,x_i} \right) \right) \\
&= \left(\frac{1}{\lambda_1} \bigvee_{i=1}^8 a_{1,1,x_i} \right) \vee \left(\frac{1}{\lambda_2} \bigvee_{i=9}^{16} a_{1,1,x_i} \right) \\
&\quad + \left(\frac{1}{\lambda_1} \bigvee_{i=1}^8 a_{1,2,x_i} \right) \vee \left(\frac{1}{\lambda_2} \bigvee_{i=9}^{16} a_{1,2,x_i} \right) \\
&= \frac{1}{\lambda_1} \times \frac{2}{3} \vee \frac{1}{\lambda_2} \times \frac{2}{3} + \frac{1}{\lambda_1} \times \frac{2}{3} \vee \frac{1}{\lambda_2} \times \frac{2}{3} \\
&= \frac{4}{3} \left(\frac{1}{\lambda_1} \vee \frac{1}{\lambda_2} \right),
\end{aligned}$$

e substituindo em (3.50), obtém-se que

$$\begin{aligned}\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2)) &= \frac{\frac{4}{3} \left(\frac{1}{\lambda_1} \vee \frac{1}{\lambda_2} \right)}{1 + \frac{4}{3} \left(\frac{1}{\lambda_1} \vee \frac{1}{\lambda_2} \right)} - \frac{1}{2} \left(\frac{\frac{4}{3}}{\lambda_1 + \frac{4}{3}} + \frac{\frac{4}{3}}{\lambda_2 + \frac{4}{3}} \right) \\ &= \frac{\frac{1}{\lambda_1 \wedge \lambda_2}}{\frac{3}{4} + \frac{1}{\lambda_1 \wedge \lambda_2}} - \frac{1}{2} \left(\frac{1}{\frac{3}{4}\lambda_1 + 1} + \frac{1}{\frac{3}{4}\lambda_2 + 1} \right) \\ &= \frac{1}{\frac{3}{4}(\lambda_1 \wedge \lambda_2) + 1} - \frac{1}{\frac{3}{2}\lambda_1 + 2} - \frac{1}{\frac{3}{2}\lambda_2 + 2}.\end{aligned}$$

Na proposição que se segue encontram-se duas propriedades relativas ao madograma: a primeira identifica o seu intervalo de variação, e a segunda relaciona o madograma com o coeficiente ε de Smith, definido em (1.28), avaliado nas regiões indexadas em I_1 , I_2 e $I_1 \cup I_2$.

Proposição 20. *Seja $\{Z(x)\}_{x \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório max-estável com funções de distribuição marginais univariadas iguais a F , com F Fréchet unitária. Então, tem-se que*

$$(i) \quad 0 \leq \nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2)) \leq \frac{1}{2};$$

$$(ii) \quad \nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_1)) = \frac{\varepsilon_{\mathbf{Z}_{I_1 \cup I_2}}}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1 \cup I_2}}} - \frac{1}{2} \left(\frac{\varepsilon_{\mathbf{Z}_{I_1}}}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1}}} + \frac{\varepsilon_{\mathbf{Z}_{I_2}}}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_2}}} \right);$$

para os conjuntos disjuntos $I_1 = \{i_0 + 1, \dots, i_1\}$ e $I_2 = \{i_1 + 1, \dots, i_2\}$, $0 = i_0 < i_1 < i_2$, $i_1, i_2 \in \mathbb{N}$, tais que $x_i \in \mathcal{D}$, com $i \in \{1, \dots, i_2\}$, e para $\lambda_1, \lambda_2 \in \mathbb{R}^+$.

Demonstração.

(i) É óbvio que

$$E \left[|F^{\lambda_1}(M(\mathbf{Z}_{I_1})) - F^{\lambda_2}(M(\mathbf{Z}_{I_2}))| \right] \geq 0.$$

Pela desigualdade triangular,

$$E \left[|F^{\lambda_1}(M(\mathbf{Z}_{I_1})) - F^{\lambda_2}(M(\mathbf{Z}_{I_2}))| \right] \leq E \left[|F^{\lambda_1}(M(\mathbf{Z}_{I_1}))| \right] + E \left[|F^{\lambda_2}(M(\mathbf{Z}_{I_2}))| \right].$$

Mas, como F é Fréchet unitária, tem-se que

$$F^{\lambda_1}(M(\mathbf{Z}_{I_1})) = F \left(\frac{M(\mathbf{Z}_{I_1})}{\lambda_1} \right) \quad \text{e} \quad F^{\lambda_2}(M(\mathbf{Z}_{I_2})) = F \left(\frac{M(\mathbf{Z}_{I_2})}{\lambda_2} \right),$$

para $\lambda_1, \lambda_2 \in \mathbb{R}^+$. E, pelo Teorema da Transformação Uniformizante 3, $F \left(\frac{M(\mathbf{Z}_{I_1})}{\lambda_1} \right), F \left(\frac{M(\mathbf{Z}_{I_2})}{\lambda_2} \right) \sim \mathcal{U}((0, 1))$. Então,

$$E \left[\left| F \left(\frac{M(\mathbf{Z}_{I_1})}{\lambda_1} \right) \right| \right] = E \left[F \left(\frac{M(\mathbf{Z}_{I_1})}{\lambda_1} \right) \right] = \frac{1}{2}.$$

e

$$E \left[\left| F \left(\frac{M(\mathbf{Z}_{I_2})}{\lambda_2} \right) \right| \right] = E \left[F \left(\frac{M(\mathbf{Z}_{I_2})}{\lambda_2} \right) \right] = \frac{1}{2}.$$

Portanto,

$$E \left[|F^{\lambda_1}(M(\mathbf{Z}_{I_1})) - F^{\lambda_2}(M(\mathbf{Z}_{I_2}))| \right] \leq 1$$

e, por conseguinte,

$$0 \leq \nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2)) \leq \frac{1}{2}.$$

(ii) Pela relação (3.50) da Proposição 19,

$$\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2)) = \frac{\varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(\lambda_1, \lambda_2)}{1 + \varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(\lambda_1, \lambda_2)} - \frac{1}{2} \left(\frac{\varepsilon_{\mathbf{Z}_{I_1}}(1)}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1}}(1)} + \frac{\varepsilon_{\mathbf{Z}_{I_2}}(1)}{\lambda_2 + \varepsilon_{\mathbf{Z}_{I_2}}(1)} \right),$$

então, fazendo $\lambda_2 = \lambda_1$, vem que

$$\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_1)) = \frac{\varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(\lambda_1, \lambda_1)}{1 + \varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(\lambda_1, \lambda_1)} - \frac{1}{2} \left(\frac{\varepsilon_{\mathbf{Z}_{I_1}}(1)}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1}}(1)} + \frac{\varepsilon_{\mathbf{Z}_{I_2}}(1)}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_2}}(1)} \right).$$

Atendendo a que a função coeficiente extremal regional coincide com a medida expoente, por (1.42), então, pela alínea (i) da Proposição 4,

$$\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2)) = \frac{\varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(1, 1)}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(1, 1)} - \frac{1}{2} \left(\frac{\varepsilon_{\mathbf{Z}_{I_1}}(1)}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1}}(1)} + \frac{\varepsilon_{\mathbf{Z}_{I_2}}(1)}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_2}}(1)} \right).$$

E, como $\varepsilon_{\mathbf{Z}_{I_1 \cup I_2}} = \varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(1, 1)$, $\varepsilon_{\mathbf{Z}_{I_1}} = \varepsilon_{\mathbf{Z}_{I_1}}(1)$ e $\varepsilon_{\mathbf{Z}_{I_2}} = \varepsilon_{\mathbf{Z}_{I_2}}(1)$,

$$\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2)) = \frac{\varepsilon_{\mathbf{Z}_{I_1 \cup I_2}}}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1 \cup I_2}}} - \frac{1}{2} \left(\frac{\varepsilon_{\mathbf{Z}_{I_1}}}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1}}} + \frac{\varepsilon_{\mathbf{Z}_{I_2}}}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_2}}} \right),$$

onde $\varepsilon_{\mathbf{Z}_{I_1}}$, $\varepsilon_{\mathbf{Z}_{I_2}}$ e $\varepsilon_{\mathbf{Z}_{I_1 \cup I_2}}$ denotam o coeficiente de Smith, definido em (1.28), avaliado nas regiões de localizações indexadas em I_1 , I_2 e $I_1 \cup I_2$, respetivamente. \square

Observação 3. Da identidade (3.50) da Proposição 19 constata-se que o madograma generalizado permite avaliar a dependência entre os vetores indexados nas regiões de localizações com índices I_1 e I_2 , através da função $\varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(\cdot)$, bem como dentro de cada uma das regiões, por meio das funções $\varepsilon_{\mathbf{Z}_{I_1}}(\cdot)$ e $\varepsilon_{\mathbf{Z}_{I_2}}(\cdot)$.

1. No caso particular em que \mathbf{Z}_{I_1} e \mathbf{Z}_{I_2} são independentes, então, de acordo com (1.45) — uma vez que o coeficiente extremal de Smith $\varepsilon_{\mathbf{Z}_{I_1 \cup I_2}}$ coincide com a função coeficiente extremal $\varepsilon_{\mathbf{Z}_{I_1}, \mathbf{Z}_{I_2}}(1, 1)$ —,

$$\varepsilon_{\mathbf{Z}_{I_1 \cup I_2}} = \varepsilon_{\mathbf{Z}_{I_1}} + \varepsilon_{\mathbf{Z}_{I_2}},$$

Assim, substituindo na expressão da alínea (ii) da Proposição 20, obtém-se:

$$\begin{aligned} \nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_1)) &= \frac{\varepsilon_{\mathbf{Z}_{I_1}} + \varepsilon_{\mathbf{Z}_{I_2}}}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1}} + \varepsilon_{\mathbf{Z}_{I_2}}} \\ &\quad - \frac{1}{2} \left(\frac{\varepsilon_{\mathbf{Z}_{I_1}}}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1}}} + \frac{\varepsilon_{\mathbf{Z}_{I_2}}}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_2}}} \right). \end{aligned}$$

2. No caso em \mathbf{Z}_{I_1} e \mathbf{Z}_{I_2} são totalmente dependentes, então, por (1.45),

$$\varepsilon_{\mathbf{Z}_{I_1 \cup I_2}} = \varepsilon_{\mathbf{Z}_{I_1}} \vee \varepsilon_{\mathbf{Z}_{I_2}}.$$

Substituindo na expressão da alínea (ii) da Proposição 20, tem-se:

$$\begin{aligned} \nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_1)) &= \frac{\varepsilon_{\mathbf{Z}_{I_1}} \vee \varepsilon_{\mathbf{Z}_{I_2}}}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1}} \vee \varepsilon_{\mathbf{Z}_{I_2}}} \\ &\quad - \frac{1}{2} \left(\frac{\varepsilon_{\mathbf{Z}_{I_1}}}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_1}}} + \frac{\varepsilon_{\mathbf{Z}_{I_2}}}{\lambda_1 + \varepsilon_{\mathbf{Z}_{I_2}}} \right). \end{aligned}$$

Estimação do madograma generalizado

Seja $(Z^{(t)}(\mathbf{x}_1), \dots, Z^{(t)}(\mathbf{x}_{i_1}), Z^{(t)}(\mathbf{x}_{i_1+1}), \dots, Z^{(t)}(\mathbf{x}_{i_2}))$, $t \in \{1, \dots, T\}$, uma amostra aleatória de $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_{i_1}), Z(\mathbf{x}_{i_1+1}), \dots, Z(\mathbf{x}_{i_2}))$ tal que as funções de distribuição das margens $Z^{(t)}(\mathbf{x}_i)$ são dadas por $F_{Z(\mathbf{x}_i)}$, $i \in \{1, \dots, i_2\}$. Repetindo o processo de transformação das margens em (3.24), agora para $Z^{(t)}(\mathbf{x}_i)$, com $i \in \{1, \dots, i_2\}$ e $t \in \{1, \dots, T\}$, de forma a obter distribuições marginais univariadas Fréchet unitárias, obtém-se o seguinte estimador fortemente consistente e assintoticamente normal (Fonseca (2012) [19]):

$$\hat{\nu}_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2)) = \frac{1}{2T} \sum_{t=1}^T \left| \bigvee_{i=1}^{i_1} F_{Z(\mathbf{x}_i)}^{\lambda_1}(Z^{(t)}(\mathbf{x}_i)) - \bigvee_{i=i_1+1}^{i_2} F_{Z(\mathbf{x}_i)}^{\lambda_2}(Z^{(t)}(\mathbf{x}_i)) \right|, \quad (3.55)$$

para $\lambda_1, \lambda_2 \in \mathbb{R}^+$, $I_1 = \{1, \dots, i_1\}$ e $I_2 = \{i_1 + 1, \dots, i_2\}$. No caso em que as funções de distribuição $F_{Z(\mathbf{x}_i)}$, $i \in \{1, \dots, i_2\}$, são desconhecidas, substituem-se pelas respetivas f.d. empíricas corrigidas.

Exemplo 17. Recorrendo novamente ao conjunto de dados **rainfall**, seleccionámos duas regiões montanhosas distintas. A região indexada em I_1 integra as estações 20, 30 e 49, situadas numa zona montanhosa interior. Por sua vez, a região com índices em I_2 é constituída pelas estações 53 e 75, localizadas numa área de montanha mais a leste. Seguindo a notação introduzida, consideramos $I_1 = \{1, 2, 3\}$ tal que $\mathbf{x}_1 = \mathbf{l}_{20}$, $\mathbf{x}_2 = \mathbf{l}_{30}$ e $\mathbf{x}_3 = \mathbf{l}_{49}$; e $I_2 = \{4, 5\}$, com $\mathbf{x}_4 = \mathbf{l}_{53}$ e $\mathbf{x}_5 = \mathbf{l}_{75}$.

Na Tabela 3.3 encontram-se as latitudes, longitudes e altitudes das localizações seleccionadas.

Tabela 3.3: Longitudes, latitudes e altitudes das localizações selecionadas no Exemplo 17.

Localização	Longitude (m)	Latitude (m)	Altitude (m)
l_{20}	699.790	220.375	910
l_{30}	712.090	218.385	908
l_{49}	690.700	214.450	805
l_{53}	744.200	234.920	2502
l_{75}	746.130	229.570	1000

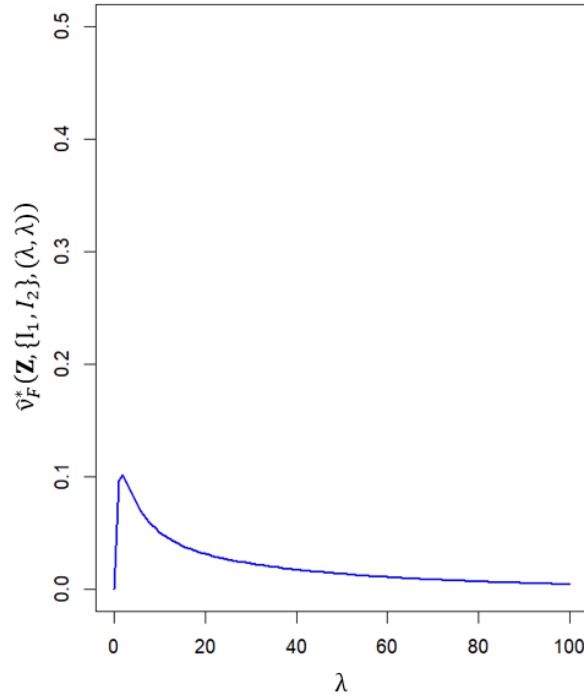


Figura 3.12: Estimativas de $\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda, \lambda))$, com $\lambda \in (0, 100)$, para o Exemplo 17.

A motivação para esta escolha reside na intenção de averiguar se a ocorrência de precipitações elevadas numa das regiões montanhosas está associada à ocorrência de precipitações elevadas na outra, não obstante as diferenças topográficas entre ambas.

A Figura 3.12 apresenta as estimativas de $\nu_F^(\mathbf{Z}, \{I_1, I_2\}, (\lambda, \lambda))$, para $\lambda \in (0, 100)$, ou seja, atribuindo igual ponderação às duas regiões. Em particular, para $\lambda = 1$, obtém-se*

$$\hat{\nu}_F^*(\mathbf{Z}, \{I_1, I_2\}, (1, 1)) \simeq 0.0966.$$

O valor máximo, 0.1026, é atingido em $\lambda \simeq 1.6237$.

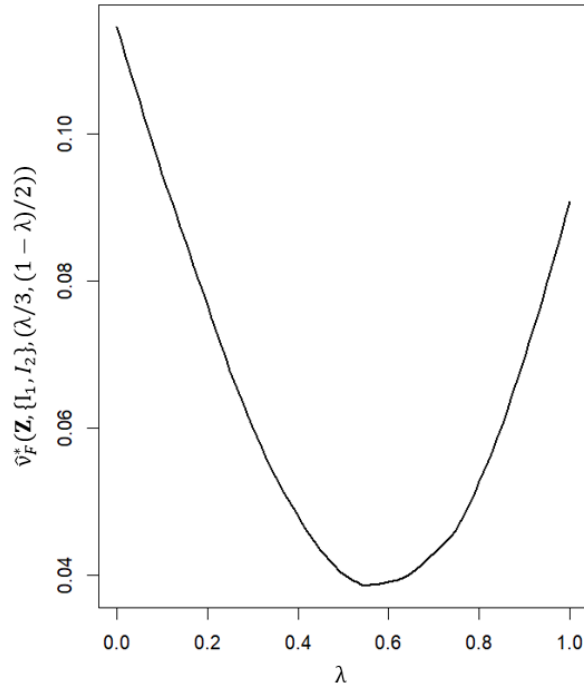


Figura 3.13: Estimativas de $\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda/3, (1-\lambda)/2))$, com $\lambda \in (0, 1)$, para o Exemplo 17.

Seguindo a alínea (iii) da Observação 2, i.e., fazendo $\lambda_1 = \frac{\lambda}{3}$ e $\lambda_2 = \frac{1-\lambda}{2}$, com $\lambda \in (0, 1)$, obtêm-se as estimativas apresentadas na Figura 3.13. Para $\lambda = 0.5$,

$$\hat{\nu}_F^*(\mathbf{Z}, \{I_1, I_2\}, (0.5/3, 0.5/2)) \simeq 0.0402.$$

O mínimo de $\hat{\nu}_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda/3, (1-\lambda)/2))$, que é aproximadamente igual a 0.0386, é atingido em $\lambda \simeq 0.5444$.

As estimativas obtidas — todas muito próximas do limite inferior teórico 0 correspondente ao caso de dependência total — sugerem uma forte dependência entre níveis elevados de precipitação nas duas regiões. Isto aponta para a possibilidade de que precipitações intensas em ambas as zonas possam, com alguma frequência, ocorrer de forma conjunta. Embora não se explorem aqui os mecanismos meteorológicos em causa, esta forte dependência pode estar associada a sistemas atmosféricos que afetam simultaneamente as duas áreas montanhosas.

A análise é complementada pelas Figuras 3.14 e 3.15, que representam a variação de $\hat{\nu}_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2))$ para diferentes combinações dos parâmetros λ_1 e λ_2 . Na Figura 3.14, λ_1 varia continuamente no intervalo $(0, 100)$, enquanto λ_2 assume os valores fixos 0.5, 1, 8, 50 e 100. A Figura 3.15 apresenta o caso em que ambos os parâmetros variam continuamente em $(0, 100)$.

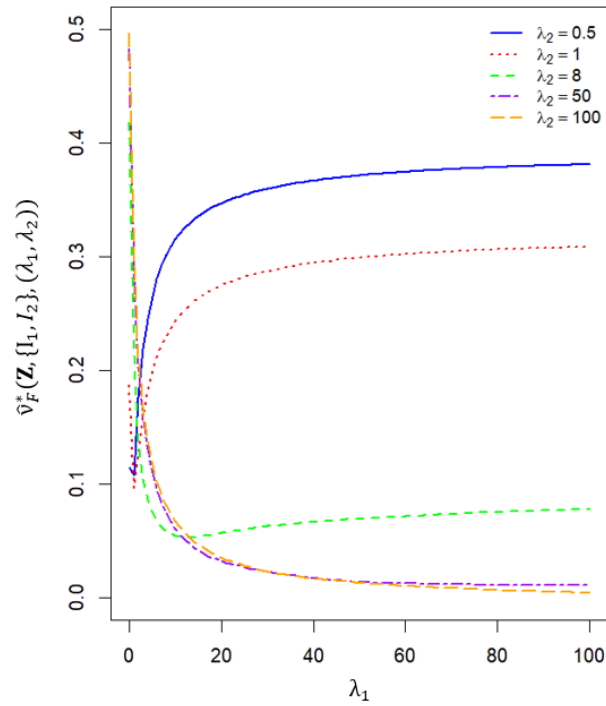


Figura 3.14: Estimativas de $\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2))$, com $\lambda_1 \in (0, 100)$ e $\lambda_2 \in \{0.5, 1, 8, 50, 100\}$, para o Exemplo 17.

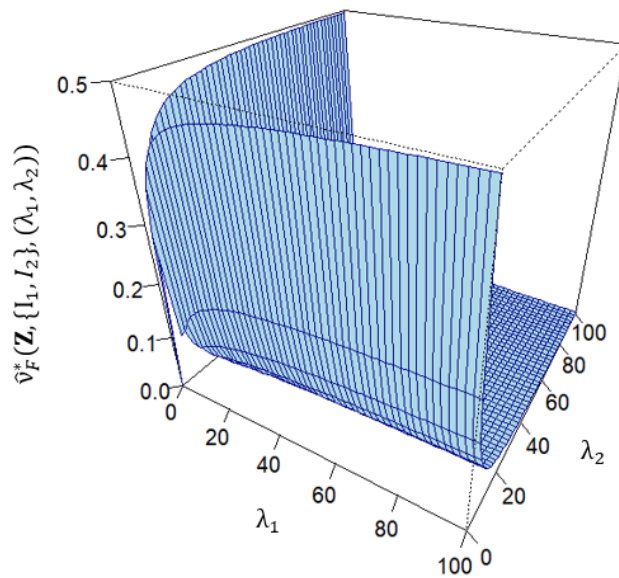


Figura 3.15: Estimativas de $\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2))$, com $\lambda_1, \lambda_2 \in (0, 100)$, para o Exemplo 17.

Madograma generalizado multiregional

Outro madograma que permite quantificar a dependência multivariada de um campo max-estável entre variáveis indexadas em mais do que duas regiões disjuntas de localizações é o introduzido por Ferreira (2013) [15], que apresentamos de seguida.

Definição 25. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório max-estável simples com funções de distribuição marginais univariadas iguais a F . Para cada vetor aleatório $\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$, para cada partição $\mathcal{I} = \{I_1 = \{i_0+1, \dots, i_1\}, I_2 = \{i_1+1, \dots, i_2\}, \dots, I_p = \{i_{p-1}+1, \dots, i_p = k\}\}$ de $\{1, \dots, k\}$, com $0 = i_0 < i_1 < i_2 < \dots < i_{p-1} < i_p$ e $1 \leq p \leq k$, e para cada $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p) \in (0, +\infty)^p$, o madograma generalizado p -regional é dado por*

$$\nu_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda}) := E \left[\prod_{j=1}^p F^{\lambda_j}(M(\mathbf{Z}_{I_j})) - \prod_{j=1}^p F^{\lambda_j}(M(\mathbf{Z}_{I_j})) \right], \quad (3.56)$$

onde $M(\mathbf{Z}_{I_j}) = \bigvee_{i \in I_j} Z(\mathbf{x}_i)$, para $j \in \{1, \dots, p\}$.

Observação 4. *A constante λ_j representa o peso atribuído à variável $F(M(\mathbf{Z}_{I_j}))$, i.e., o peso atribuído à região de localizações indexadas em I_j , com $j \in \{1, \dots, p\}$.*

É importante frisar que (3.56) pode ser reescrito nas formas

$$\nu_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda}) = E \left[\prod_{i,j \in \{1, \dots, p\}} \left| F^{\lambda_i}(M(\mathbf{Z}_{I_i})) - F^{\lambda_j}(M(\mathbf{Z}_{I_j})) \right| \right] \quad (3.57)$$

e

$$\nu_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda}) = E \left[\prod_{j=1}^p \prod_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) - \prod_{j=1}^p \prod_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \right]. \quad (3.58)$$

A equação (3.57) evidencia que o madograma generalizado p -regional corresponde ao dobro do madograma generalizado definido em (3.48), se $p = 2$ e $k = i_2$, i.e.,

$$\nu_F(\mathbf{Z}, \mathcal{I} = \{I_1 = \{1, \dots, i_1\}, I_2 = \{i_1 + 1, \dots, i_2\}\}, \boldsymbol{\lambda} = (\lambda_1, \lambda_2)) = 2\nu_F^*(\mathbf{Z}, \{I_1, I_2\}, (\lambda_1, \lambda_2)). \quad (3.59)$$

Observação 5. *Sendo o λ -madograma um caso particular do madograma generalizado, também será do madograma generalizado p -regional. Se se pretender estender a abordagem do λ -madograma a p regiões e, simultaneamente, garantir que não se privilegia nenhuma região pelo seu número de localizações, pode-se considerar $\lambda_1 = \frac{\alpha_1}{|I_1|}, \dots, \lambda_p = \frac{\alpha_p}{|I_p|}$, com $\alpha_1, \dots, \alpha_p \in (0, 1)$ e $\alpha_1 + \dots + \alpha_p = 1$.*

A equação (3.58) induz o estimador fortemente consistente e assintoticamente normal que apresentaremos adiante. Além disso, possibilita a demonstração de propriedades fundamentais do madograma como as apresentadas na próxima proposição, onde são estabelecidas

identidades envolvendo as funções de 1.2.3 e o coeficiente extremal de Smith, que, por sua vez, facilitam o cálculo do madograma.

Proposição 21. *Seja $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\mathcal{D} \subseteq \mathbb{R}^d$, um campo aleatório max-estável simples com funções de distribuição marginais univariadas iguais a F . Para cada vetor aleatório $\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$, para cada partição $\mathcal{I} = \{I_1 = \{i_0+1, \dots, i_1\}, I_2 = \{i_1+1, \dots, i_2\}, \dots, I_p = \{i_{p-1}+1, \dots, i_p = k\}\}$ de $\{1, \dots, k\}$, com $0 = i_0 < i_1 < i_2 < \dots < i_{p-1} < i_p$ e $1 \leq p \leq k$, e para cada $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p) \in (0, +\infty)^p$, o madograma generalizado p -regional verifica*

$$\nu_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda}) = m_{\mathbf{Z}_{I_j, j \in \{1, \dots, p\}}}(\lambda_j, j \in \{1, \dots, p\}) - \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} m_{\mathbf{Z}_{I_j, j \in S}}(\lambda_j, j \in S), \quad (3.60)$$

com $m_{\mathbf{Z}_{I_j, j \in S}}(\lambda_j, j \in S)$, para $\emptyset \neq S \subseteq \{1, \dots, p\}$, a função definida em (1.39), e

$$\nu_F(\mathbf{Z}, \mathcal{I}, \mathbf{1}) = \frac{\varepsilon_{\mathbf{Z}}}{1 + \varepsilon_{\mathbf{Z}}} - \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} \frac{\varepsilon_{\mathbf{Z}_{\cup_{j \in S} I_j}}}{1 + \varepsilon_{\mathbf{Z}_{\cup_{j \in S} I_j}}}, \quad (3.61)$$

com ε o coeficiente extremal de Smith definido em (1.28).

Demonstração. Da equação (3.58) deduz-se que

$$\nu_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda}) = E \left[\bigvee_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \right] - E \left[\bigwedge_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \right]. \quad (3.62)$$

Ora, por (1.39),

$$E \left[\bigvee_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \right] = m_{\mathbf{Z}_{I_j, j \in \{1, \dots, p\}}}(\lambda_j, j \in \{1, \dots, p\}). \quad (3.63)$$

Por outro lado, pela Fórmula de Poincaré ²,

$$\begin{aligned} P \left(\bigwedge_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \leq u \right) &= P \left(- \bigvee_{j=1}^p - \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \leq u \right) \\ &= P \left(\bigvee_{j=1}^p - \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \geq -u \right) \end{aligned}$$

²Fórmula de Poincaré:

Para os acontecimentos A_1, \dots, A_n , a igualdade

$$P \left(\bigcup_{j=1}^n A_j \right) = \sum_{\emptyset \neq S \subseteq \{1, \dots, n\}} (-1)^{|S|+1} P \left(\bigcap_{j \in S} A_j \right)$$

é válida.

$$\begin{aligned}
&= P \left(\bigcup_{j=1}^p \left\{ - \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \geq -u \right\} \right) \\
&= P \left(\bigcup_{j=1}^p \left\{ \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \leq u \right\} \right) \\
&= \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} P \left(\bigcap_{j \in S} \left\{ \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \leq u \right\} \right) \\
&= \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} P \left(\bigcap_{j \in S} \left\{ \bigcap_{i \in I_j} \left\{ F^{\lambda_j}(Z(\mathbf{x}_i)) \leq u \right\} \right\} \right), \quad 0 < u < 1,
\end{aligned}$$

donde, tendo em conta

$$P \left(\bigcap_{j \in S} \left\{ \bigcap_{i \in I_j} \left\{ F^{\lambda_j}(Z(\mathbf{x}_i)) \leq u \right\} \right\} \right) = P \left(\bigvee_{j \in S} \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \leq u \right)$$

e (3.63),

$$E \left[\bigwedge_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \right] = \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} m_{\mathbf{Z}_{I_j}, j \in S}(\lambda_j, j \in S). \quad (3.64)$$

Aplicando (3.63) e (3.64) em (3.62), obtém-se (3.60):

$$\nu_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda}) = m_{\mathbf{Z}_{I_j}, j \in \{1, \dots, p\}}(\lambda_j, j \in \{1, \dots, p\}) - \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} m_{\mathbf{Z}_{I_j}, j \in S}(\lambda_j, j \in S).$$

Atendendo a que

$$m_{\mathbf{Z}_{I_j}, j \in S}(\lambda_j, j \in S) = \frac{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_j \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_j \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right)}{1 + V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_j \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_j \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right)},$$

por (1.40), a equação anterior pode ser reescrita na forma

$$\begin{aligned}
\nu_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda}) &= \frac{V_{\mathbf{Z}} \left(\sum_{j=1}^p \lambda_j \delta_1(I_j), \dots, \sum_{j=1}^p \lambda_j \delta_k(I_j) \right)}{1 + V_{\mathbf{Z}} \left(\sum_{j=1}^p \lambda_j \delta_1(I_j), \dots, \sum_{j=1}^p \lambda_j \delta_k(I_j) \right)} \\
&\quad - \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} \frac{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_j \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_j \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right)}{1 + V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_j \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_j \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right)},
\end{aligned}$$

e pela forma como a partição \mathcal{I} está definida,

$$\begin{aligned} \nu_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda}) &= \frac{V_{\mathbf{Z}}(\lambda_1, \dots, \lambda_p)}{1 + V_{\mathbf{Z}}(\lambda_1, \dots, \lambda_p)} \\ &- \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} \frac{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_j \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_j \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right)}{1 + V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}} \left(\sum_{j \in S} \lambda_j \delta_{\alpha}(\bigcup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_j \delta_{\omega}(\bigcup_{j \in S} I_j)(I_j) \right)}. \end{aligned} \quad (3.65)$$

Em particular, quando $\boldsymbol{\lambda} = \mathbf{1}$, tem-se

$$\nu_F(\mathbf{Z}, \mathcal{I}, \mathbf{1}) = \frac{V_{\mathbf{Z}}(1, \dots, 1)}{1 + V_{\mathbf{Z}}(1, \dots, 1)} - \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} \frac{V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}}(1, \dots, 1)}{1 + V_{\mathbf{Z}_{\bigcup_{j \in S} I_j}}(1, \dots, 1)}.$$

Pela relação entre o coeficiente extremal de Smith e a medida expoente em (1.29), deriva-se (3.61):

$$\nu_F(\mathbf{Z}, \mathcal{I}, \mathbf{1}) = \frac{\varepsilon_{\mathbf{Z}}}{1 + \varepsilon_{\mathbf{Z}}} - \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} \frac{\varepsilon_{\mathbf{Z}_{\bigcup_{j \in S} I_j}}}{1 + \varepsilon_{\mathbf{Z}_{\bigcup_{j \in S} I_j}}}.$$

□

Observação 6. *Supondo que estamos nas condições da proposição anterior, apliquemo-la para determinar a expressão do madograma quando os elementos da partição \mathcal{I} são da forma $I_j = \{j\}, j = 1, \dots, p = k$. Nesse caso, usando (3.65),*

$$\begin{aligned} \nu_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda}) &= \frac{V_{\mathbf{Z}}(\lambda_1, \dots, \lambda_k)}{1 + V_{\mathbf{Z}}(\lambda_1, \dots, \lambda_k)} \\ &- \sum_{\emptyset \neq S \subseteq \{1, \dots, k\}} (-1)^{|S|+1} \frac{V_{\mathbf{Z}_{\bigcup_{j \in S} \{j\}}} \left(\sum_{j \in S} \lambda_j \delta_{\alpha}(\bigcup_{j \in S} \{j\})(\{j\}), \dots, \sum_{j \in S} \lambda_j \delta_{\omega}(\bigcup_{j \in S} \{j\})(\{j\}) \right)}{1 + V_{\mathbf{Z}_{\bigcup_{j \in S} \{j\}}} \left(\sum_{j \in S} \lambda_j \delta_{\alpha}(\bigcup_{j \in S} \{j\})(\{j\}), \dots, \sum_{j \in S} \lambda_j \delta_{\omega}(\bigcup_{j \in S} \{j\})(\{j\}) \right)} \\ &= \frac{V_{\mathbf{Z}}(\lambda_1, \dots, \lambda_k)}{1 + V_{\mathbf{Z}}(\lambda_1, \dots, \lambda_k)} - \sum_{\emptyset \neq S \subseteq \{1, \dots, k\}} (-1)^{|S|+1} \frac{V_{\mathbf{Z}_S}(\lambda_{\alpha(S)}, \dots, \lambda_{\omega(S)})}{1 + V_{\mathbf{Z}_S}(\lambda_{\alpha(S)}, \dots, \lambda_{\omega(S)})}. \end{aligned} \quad (3.66)$$

Em particular, quando $\boldsymbol{\lambda} = \mathbf{1}$, de (3.61) tem-se

$$\nu_F(\mathbf{Z}, \mathcal{I}, \mathbf{1}) = \frac{\varepsilon_{\mathbf{Z}}}{1 + \varepsilon_{\mathbf{Z}}} - \sum_{\emptyset \neq S \subseteq \{1, \dots, k\}} (-1)^{|S|+1} \frac{\varepsilon_{\mathbf{Z}_S}}{1 + \varepsilon_{\mathbf{Z}_S}}. \quad (3.67)$$

Considere-se agora $k = 2$ e que o campo $\{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ é estacionário.

(i) No caso em que $\lambda_1 = 1 - \lambda_2 = \lambda$, usando a expressão obtida em (3.66), tem-se que

$$\begin{aligned}
\frac{1}{2}\nu_F(\mathbf{Z}, \mathcal{I}, \lambda) &= \frac{1}{2} \left(\frac{V_{\mathbf{Z}}(\lambda, 1 - \lambda)}{1 + V_{\mathbf{Z}}(\lambda, 1 - \lambda)} - \sum_{\emptyset \neq S \subseteq \{1,2\}} (-1)^{|S|+1} \frac{V_{\mathbf{Z}_S}(\lambda_{\alpha(S)}, \lambda_{\omega(S)})}{1 + V_{\mathbf{Z}_S}(\lambda_{\alpha(S)}, \lambda_{\omega(S)})} \right) \\
&= \frac{1}{2} \frac{V_{\mathbf{Z}}(\lambda, 1 - \lambda)}{1 + V_{\mathbf{Z}}(\lambda, 1 - \lambda)} - \frac{1}{2} \left(\frac{V_{\mathbf{Z}_{\{1\}}}(\lambda)}{1 + V_{\mathbf{Z}_{\{1\}}}(\lambda)} + \frac{V_{\mathbf{Z}_{\{2\}}}(1 - \lambda)}{1 + V_{\mathbf{Z}_{\{2\}}}(1 - \lambda)} - \frac{V_{\mathbf{Z}}(\lambda, 1 - \lambda)}{1 + V_{\mathbf{Z}}(\lambda, 1 - \lambda)} \right) \\
&\stackrel{4}{=} \frac{V_{\mathbf{Z}}(\lambda, 1 - \lambda)}{1 + V_{\mathbf{Z}}(\lambda, 1 - \lambda)} - \frac{1}{2} \left(\frac{1/\lambda}{1 + 1/\lambda} + \frac{1/(1 - \lambda)}{1 + 1/(1 - \lambda)} \right) \\
&= \frac{V_{\mathbf{Z}}(\lambda, 1 - \lambda)}{1 + V_{\mathbf{Z}}(\lambda, 1 - \lambda)} - \frac{1}{2} \left(\frac{1}{\lambda + 1} + \frac{1}{1 - \lambda + 1} \right) \\
&= \frac{V_{\mathbf{Z}}(\lambda, 1 - \lambda)}{V_{\mathbf{Z}}(\lambda, 1 - \lambda) + 1} - \frac{1}{2} \frac{2 - \lambda + 1 + \lambda}{(2 - \lambda)(1 + \lambda)} \\
&= \frac{V_{\mathbf{Z}}(\lambda, 1 - \lambda)}{V_{\mathbf{Z}}(\lambda, 1 - \lambda) + 1} - \frac{3}{2(2 - \lambda)(1 + \lambda)},
\end{aligned}$$

que é precisamente a equação (3.29) da Proposição 15, já que $V_{\mathbf{Z}}(\lambda, 1 - \lambda) = \varepsilon_{\mathbf{Z}}(\lambda, 1 - \lambda)$, segundo (1.42).

(ii) Já no caso em que $\lambda_1 = \lambda_2 = 1$, usa-se a expressão (3.67) e obtém-se

$$\begin{aligned}
\frac{1}{2}\nu_F(\mathbf{Z}, \mathcal{I}, \mathbf{1}) &= \frac{1}{2} \frac{\varepsilon_{\mathbf{Z}}}{1 + \varepsilon_{\mathbf{Z}}} - \frac{1}{2} \sum_{\emptyset \neq S \subseteq \{1,2\}} (-1)^{|S|+1} \frac{\varepsilon_{\mathbf{Z}_S}}{1 + \varepsilon_{\mathbf{Z}_S}} \\
&= \frac{1}{2} \frac{\varepsilon_{\mathbf{Z}}}{1 + \varepsilon_{\mathbf{Z}}} - \frac{1}{2} \left(\frac{\varepsilon_{\mathbf{Z}_{\{1\}}}}{1 + \varepsilon_{\mathbf{Z}_{\{1\}}}} + \frac{\varepsilon_{\mathbf{Z}_{\{2\}}}}{1 + \varepsilon_{\mathbf{Z}_{\{2\}}}} - \frac{\varepsilon_{\mathbf{Z}}}{1 + \varepsilon_{\mathbf{Z}}} \right) \\
&= \frac{1}{2} \frac{\varepsilon_{\mathbf{Z}}}{1 + \varepsilon_{\mathbf{Z}}} - \frac{1}{2} \left(\frac{1}{2} + \frac{1}{2} - \frac{\varepsilon_{\mathbf{Z}}}{1 + \varepsilon_{\mathbf{Z}}} \right) \\
&= \frac{\varepsilon_{\mathbf{Z}}}{1 + \varepsilon_{\mathbf{Z}}} - \frac{1}{2},
\end{aligned}$$

que coincide com a equação (3.20) da Proposição 13, uma vez que, dada a estacionariedade, $\varepsilon_{\mathbf{Z}} = \varepsilon(\mathbf{h})$, sempre que $\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{h}$.

Exemplo 18. Apliquemos agora a Proposição 21 ao campo M_4 introduzido no Exemplo 14 para calcular o valor do madograma para o vetor $\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_{16}))$, a partição $\mathcal{I} = \{I_1 = \{1, 2, 3, 4, 5, 6\}, I_2 = \{7, 8, 9, 10\}, I_3 = \{11, 12, 13, 14, 15, 16\}\}$ e $\lambda = \{\lambda_1, \lambda_2, \lambda_3\}$.

As coordenadas dos pontos \mathbf{x}_i , com $i \in \{1, \dots, 16\}$, são apresentadas na Figura 3.16.

Pela equação (3.65),

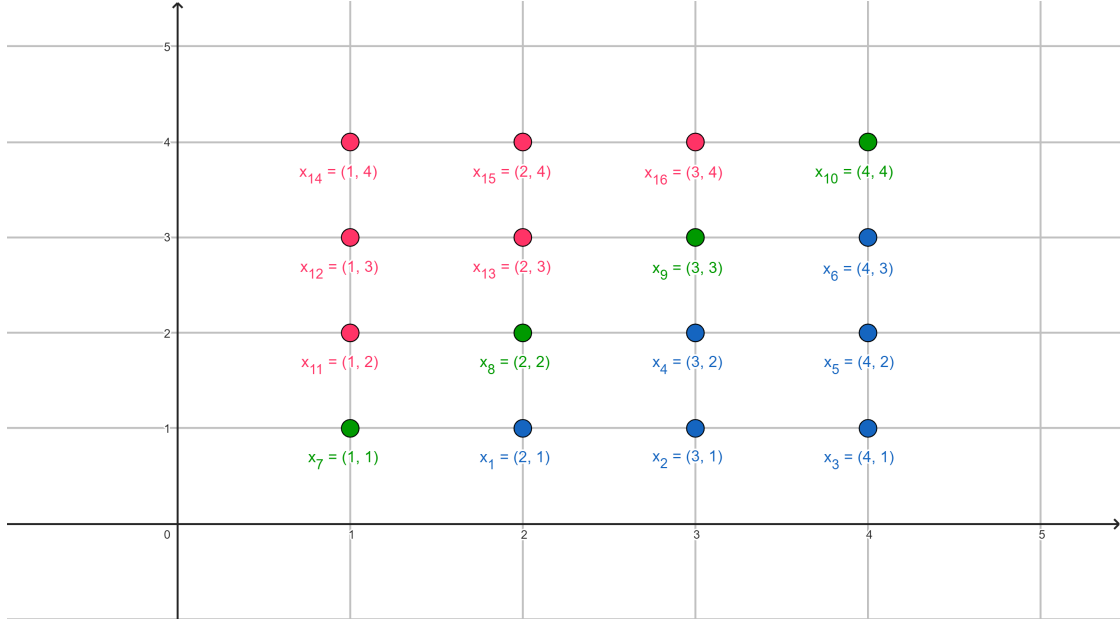


Figura 3.16: Coordenadas dos pontos indexados na partição \mathcal{I} do Exemplo 18.

$$\begin{aligned}
\nu_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda}) &= \frac{V_{\mathbf{Z}}(\lambda_1, \dots, \lambda_3)}{1 + V_{\mathbf{Z}}(\lambda_1, \dots, \lambda_3)} \\
&- \sum_{\emptyset \neq S \subseteq \{1,2,3\}} (-1)^{|S|+1} \frac{V_{\mathbf{Z} \cup_{j \in S} I_j} \left(\sum_{j \in S} \lambda_j \delta_{\alpha}(\cup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_j \delta_{\omega}(\cup_{j \in S} I_j)(I_j) \right)}{1 + V_{\mathbf{Z} \cup_{j \in S} I_j} \left(\sum_{j \in S} \lambda_j \delta_{\alpha}(\cup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_j \delta_{\omega}(\cup_{j \in S} I_j)(I_j) \right)} \\
&= 1 - \frac{1}{1 + V_{\mathbf{Z}}(\lambda_1, \dots, \lambda_3)} \\
&- \sum_{\emptyset \neq S \subseteq \{1,2,3\}} (-1)^{|S|+1} \left(1 - \frac{1}{1 + V_{\mathbf{Z} \cup_{j \in S} I_j} \left(\sum_{j \in S} \lambda_j \delta_{\alpha}(\cup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_j \delta_{\omega}(\cup_{j \in S} I_j)(I_j) \right)} \right) \\
&= \sum_{\emptyset \neq S \subseteq \{1,2,3\}} (-1)^{|S|+1} \frac{1}{1 + V_{\mathbf{Z} \cup_{j \in S} I_j} \left(\sum_{j \in S} \lambda_j \delta_{\alpha}(\cup_{j \in S} I_j)(I_j), \dots, \sum_{j \in S} \lambda_j \delta_{\omega}(\cup_{j \in S} I_j)(I_j) \right)} \\
&- \frac{1}{1 + V_{\mathbf{Z}}(\lambda_1, \dots, \lambda_3)},
\end{aligned}$$

onde a última igualdade segue de

$$\sum_{\emptyset \neq S \subseteq \{1, \dots, k\}} (-1)^{|S|+1} = 1 \tag{3.68}$$

Por (3.44),

$$\begin{aligned}
V_{\mathbf{Z} \cup_{j \in S} I_j} \left(\sum_{j \in S} \lambda_j \delta_{\alpha(\cup_{j \in S} I_j)}(I_j), \dots, \sum_{j \in S} \lambda_j \delta_{\omega(\cup_{j \in S} I_j)}(I_j) \right) &= \sum_{l \in \mathbb{N}} \sum_{m \in \mathbb{Z}} \prod_{i=\alpha(\cup_{j \in S} I_j)}^{\omega(\cup_{j \in S} I_j)} a_{l,m,\mathbf{x}_i} \left[\sum_{j \in S} \lambda_j \delta_i(I_j) \right]^{-1} \\
&= \sum_{l \in \mathbb{N}} \sum_{m \in \mathbb{Z}} \prod_{i=\alpha(\cup_{j \in S} I_j)}^{\omega(\cup_{j \in S} I_j)} a_{l,m,\mathbf{x}_i} \lambda_j^{-1}, \quad i \in I_{j'} \\
&= \sum_{l \in \mathbb{N}} \sum_{m \in \mathbb{Z}} \prod_{j \in S} \prod_{i \in I_j} \lambda_j^{-1} a_{l,m,\mathbf{x}_i}.
\end{aligned} \tag{3.69}$$

E, por conseguinte,

$$\begin{aligned}
\nu_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda}) &= \sum_{\emptyset \neq S \subseteq \{1,2,3\}} (-1)^{|S|+1} \frac{1}{1 + \sum_{m=1}^2 \prod_{j \in S} \prod_{i \in I_j} \lambda_j^{-1} a_{l,m,\mathbf{x}_i}} - \frac{1}{1 + \sum_{m=1}^2 \prod_{j=1}^3 \prod_{i \in I_j} \lambda_j^{-1} a_{l,m,\mathbf{x}_i}} \\
&= \sum_{\emptyset \neq S \subseteq \{1,2,3\}} (-1)^{|S|+1} \frac{1}{1 + \sum_{m=1}^2 \prod_{j \in S} \prod_{i \in I_j} \lambda_j^{-1} a_{l,m,\mathbf{x}_i}} + \frac{(-1)^{3+1}}{1 + \sum_{m=1}^2 \prod_{j=1}^3 \prod_{i \in I_j} \lambda_j^{-1} a_{l,m,\mathbf{x}_i}} \\
&\quad - \frac{1}{1 + \sum_{m=1}^2 \prod_{j=1}^3 \prod_{i \in I_j} \lambda_j^{-1} a_{l,m,\mathbf{x}_i}} \\
&= \sum_{\emptyset \neq S \subseteq \{1,2,3\}} (-1)^{|S|+1} \frac{1}{1 + \sum_{m=1}^2 \prod_{j \in S} \prod_{i \in I_j} \lambda_j^{-1} a_{l,m,\mathbf{x}_i}} \\
&= \frac{(-1)^{1+1}}{1 + \sum_{m=1}^2 \prod_{i \in I_1} \lambda_1^{-1} a_{l,m,\mathbf{x}_i}} + \frac{(-1)^{1+1}}{1 + \sum_{m=1}^2 \prod_{i \in I_2} \lambda_2^{-1} a_{l,m,\mathbf{x}_i}} + \frac{(-1)^{1+1}}{1 + \sum_{m=1}^2 \prod_{i \in I_3} \lambda_3^{-1} a_{l,m,\mathbf{x}_i}} \\
&\quad + \frac{(-1)^{2+1}}{1 + \sum_{m=1}^2 \prod_{j \in \{1,2\}} \prod_{i \in I_j} \lambda_j^{-1} a_{l,m,\mathbf{x}_i}} + \frac{(-1)^{2+1}}{1 + \sum_{m=1}^2 \prod_{j \in \{1,3\}} \prod_{i \in I_j} \lambda_j^{-1} a_{l,m,\mathbf{x}_i}} \\
&\quad + \frac{(-1)^{2+1}}{1 + \sum_{m=1}^2 \prod_{j \in \{2,3\}} \prod_{i \in I_j} \lambda_j^{-1} a_{l,m,\mathbf{x}_i}} \\
&= \frac{1}{1 + \frac{4}{3} \lambda_1^{-1}} + \frac{1}{1 + \frac{7}{6} \lambda_2^{-1}} + \frac{1}{1 + \frac{4}{3} \lambda_3^{-1}} - \frac{1}{1 + \frac{2}{3} (\lambda_1^{-1} \vee \lambda_2^{-1})} + \frac{2}{3} \lambda_1^{-1} \vee \frac{1}{2} \lambda_2^{-1} - \frac{1}{1 + \frac{4}{3} (\lambda_1^{-1} \vee \lambda_3^{-1})} \\
&\quad - \frac{1}{1 + \frac{2}{3} (\lambda_2^{-1} \vee \lambda_3^{-1})} + \frac{1}{2} \lambda_2^{-1} + \frac{2}{3} \lambda_3^{-1}.
\end{aligned}$$

Os cálculos que fundamentam a última passagem encontram-se em Apêndice.

Na próxima proposição ressalta-se que, independentemente da partição considerada, o valor do madograma tem como extremantes os casos de dependência total e independência entre as margens do vetor do campo em que é calculado, e que, em particular, quando o vetor $\boldsymbol{\lambda}$ é o vetor constante igual a 1, um dos extremos é nulo e o outro é facilmente obtido através do coeficiente de Smith.

Proposição 22. *Seja $\mathcal{I} = \{I_1 = \{i_0 + 1, \dots, i_1\}, I_2 = \{i_1 + 1, \dots, i_2\}, \dots, I_p = \{i_{p-1} + 1, \dots, i_p = k\}\}$ uma partição arbitrária de $\{1, \dots, k\}$, tal que $0 = i_0 < i_1 < i_2 < \dots < i_{p-1} < i_p$ e $1 \leq p \leq k$.*

Considerem-se $Z = \{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$, $\hat{Z} = \{\hat{Z}(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ e $\tilde{Z} = \{\tilde{Z}(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ campos aleatórios max-estáveis simples, com funções de distribuição marginais univariadas iguais a F , tais que $\hat{Z}_{I_j}, j \in \{1, \dots, p\}$, são independentes, $\tilde{Z}_{I_j}, j \in \{1, \dots, p\}$, são totalmente dependentes e, para cada $j \in \{1, \dots, p\}$, Z_{I_j}, \hat{Z}_{I_j} e \tilde{Z}_{I_j} são identicamente distribuídos.

Então, para quaisquer vetores $\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$, $\hat{\mathbf{Z}} = (\hat{Z}(\mathbf{x}_1), \dots, \hat{Z}(\mathbf{x}_k))$ e $\tilde{\mathbf{Z}} = (\tilde{Z}(\mathbf{x}_1), \dots, \tilde{Z}(\mathbf{x}_k))$, para qualquer vetor $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p) \in (0, +\infty)^p$ e para a partição \mathcal{I} , tem-se

$$(i) \quad \nu_F(\tilde{\mathbf{Z}}, \mathcal{I}, \boldsymbol{\lambda}) \leq \nu_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda}) \leq \nu_F(\hat{\mathbf{Z}}, \mathcal{I}, \boldsymbol{\lambda});$$

$$(ii) \quad \nu_F(\tilde{\mathbf{Z}}, \mathcal{I}, \mathbf{1}) = 0;$$

$$(iii) \quad \nu_F(\hat{\mathbf{Z}}, \mathcal{I}, \mathbf{1}) = \frac{\sum_{j=1}^p \varepsilon^{\hat{Z}_{I_j}}}{\sum_{j=1}^p \varepsilon^{\hat{Z}_{I_j}} + 1} - \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} \frac{\sum_{j \in S} \varepsilon^{\hat{Z}_{I_j}}}{\sum_{j \in S} \varepsilon^{\hat{Z}_{I_j}} + 1},$$

onde ε denota o coeficiente extremal de Smith definido em (1.28).

Demonstração.

(i) Atendendo a que Z é um campo max-estável, então as distribuições marginais univariadas e multivariadas de Z são de valores extremos e, conseqüentemente, associadas, segundo Marshall e Olkin (1983) [32]. Sendo as distribuições de valores extremos max-estáveis, também as variáveis $M(\mathbf{Z}_{I_j}) = \frac{\bigvee_{i \in I_j} Z(\mathbf{x}_i)}{\lambda_j}$, $j \in \{1, \dots, p\}$, são associadas e, por conseguinte, dependentes no ortante positivo, pelo que as desigualdades

$$P \left(\bigcap_{j=1}^p \{M(\mathbf{Z}_{I_j}) \leq z_j\} \right) \geq \prod_{j=1}^p P(M(\mathbf{Z}_{I_j}) \leq z_j) \quad (3.70)$$

e

$$P \left(\bigcap_{j=1}^p \{M(\mathbf{Z}_{I_j}) > z_j\} \right) \geq \prod_{j=1}^p P(M(\mathbf{Z}_{I_j}) > z_j) \quad (3.71)$$

são válidas para $z_1, \dots, z_p \in \mathbb{R}^+$. Por outro lado, como $\left\{ \bigcap_{j=1}^p \{M(\mathbf{Z}_{I_j}) \leq z_j\} \right\} \subseteq \{M(\mathbf{Z}_{I_j}) \leq z_j\}$ e $\left\{ \bigcap_{j=1}^p \{M(\mathbf{Z}_{I_j}) > z_j\} \right\} \subseteq \{M(\mathbf{Z}_{I_j}) > z_j\}$, para qualquer $j \in \{1, \dots, p\}$, tem-se que

$$P \left(\bigcap_{j=1}^p \{M(\mathbf{Z}_{I_j}) \leq z_j\} \right) \leq P(M(\mathbf{Z}_{I_j}) \leq z_j)$$

e

$$P \left(\bigcap_{j=1}^p \{M(\mathbf{Z}_{I_j}) > z_j\} \right) \leq P(M(\mathbf{Z}_{I_j}) > z_j),$$

para qualquer $j \in \{1, \dots, p\}$. Portanto,

$$P \left(\bigcap_{j=1}^p \{M(\mathbf{Z}_{I_j}) \leq z_j\} \right) \leq \bigwedge_{j=1}^p P(M(\mathbf{Z}_{I_j}) \leq z_j) \quad (3.72)$$

e

$$P \left(\bigcap_{j=1}^p \{M(\mathbf{Z}_{I_j}) > z_j\} \right) \leq \bigwedge_{j=1}^p P(M(\mathbf{Z}_{I_j}) > z_j). \quad (3.73)$$

Juntando (3.70) com (3.72) e (3.71) com (3.73), segue-se que

$$\prod_{j=1}^p P(M(\mathbf{Z}_{I_j}) \leq z_j) \leq P \left(\bigcap_{j=1}^p \{M(\mathbf{Z}_{I_j}) \leq z_j\} \right) \leq \bigwedge_{j=1}^p P(M(\mathbf{Z}_{I_j}) \leq z_j)$$

e

$$\prod_{j=1}^p P(M(\mathbf{Z}_{I_j}) > z_j) \leq P \left(\bigcap_{j=1}^p \{M(\mathbf{Z}_{I_j}) > z_j\} \right) \leq \bigwedge_{j=1}^p P(M(\mathbf{Z}_{I_j}) > z_j).$$

Atendendo a que os limites inferior e superior correspondem, respetivamente, aos casos de independência e dependência total entre $M(\mathbf{Z}_{I_j})$, $j \in \{1, \dots, p\}$, e os campos Z , \hat{Z} e \tilde{Z} são identicamente distribuídos, então, considerando $M(\hat{\mathbf{Z}}_{I_j}) = \frac{\bigvee_{i \in I_j} \hat{Z}(\mathbf{x}_i)}{\lambda_j}$ e $M(\tilde{\mathbf{Z}}_{I_j}) = \frac{\bigvee_{i \in I_j} \tilde{Z}(\mathbf{x}_i)}{\lambda_j}$, para $j \in \{1, \dots, p\}$, os enquadramentos anteriores podem ser reescritos nas formas

$$P \left(\bigcap_{j=1}^p \{M(\hat{\mathbf{Z}}_{I_j}) \leq z_j\} \right) \leq P \left(\bigcap_{j=1}^p \{M(\mathbf{Z}_{I_j}) \leq z_j\} \right) \leq P \left(\bigcap_{j=1}^p \{M(\tilde{\mathbf{Z}}_{I_j}) \leq z_j\} \right)$$

e

$$P \left(\bigcap_{j=1}^p \{M(\hat{\mathbf{Z}}_{I_j}) > z_j\} \right) \leq P \left(\bigcap_{j=1}^p \{M(\mathbf{Z}_{I_j}) > z_j\} \right) \leq P \left(\bigcap_{j=1}^p \{M(\tilde{\mathbf{Z}}_{I_j}) > z_j\} \right).$$

Assim, por Shaked e Shanthikumar (2007) [41], tem-se

$$E \left[F \left(\bigvee_{j=1}^p M(\hat{\mathbf{Z}}_{I_j}) \right) \right] \geq E \left[F \left(\bigvee_{j=1}^p M(\mathbf{Z}_{I_j}) \right) \right] \geq E \left[F \left(\bigvee_{j=1}^p M(\tilde{\mathbf{Z}}_{I_j}) \right) \right]$$

e

$$E \left[F \left(\bigwedge_{j=1}^p M(\hat{\mathbf{Z}}_{I_j}) \right) \right] \geq E \left[F \left(\bigwedge_{j=1}^p M(\mathbf{Z}_{I_j}) \right) \right] \geq E \left[F \left(\bigwedge_{j=1}^p M(\tilde{\mathbf{Z}}_{I_j}) \right) \right].$$

Pelas definições de $M(\mathbf{Z}_{I_j})$, $M(\hat{\mathbf{Z}}_{I_j})$ e $M(\tilde{\mathbf{Z}}_{I_j})$ e por F ser Fréchet unitária, os dois últimos enquadramentos são equivalentes a

$$E \left[\bigvee_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(\hat{Z}(\mathbf{x}_i)) \right] \geq E \left[\bigvee_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \right] \geq E \left[\bigvee_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(\tilde{Z}(\mathbf{x}_i)) \right]$$

e

$$E \left[\bigwedge_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(\hat{Z}(\mathbf{x}_i)) \right] \leq E \left[\bigwedge_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \right] \leq E \left[\bigwedge_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(\tilde{Z}(\mathbf{x}_i)) \right].$$

Subtraindo as desigualdades da última linha às da penúltima, obtém-se que

$$\begin{aligned} E \left[\bigvee_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(\tilde{Z}(\mathbf{x}_i)) - \bigwedge_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(\tilde{Z}(\mathbf{x}_i)) \right] &\leq E \left[\bigvee_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) - \bigwedge_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(Z(\mathbf{x}_i)) \right] \\ &= E \left[\bigvee_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(\hat{Z}(\mathbf{x}_i)) - \bigwedge_{j=1}^p \bigvee_{i \in I_j} F^{\lambda_j}(\hat{Z}(\mathbf{x}_i)) \right], \end{aligned}$$

ou seja,

$$\nu_F(\tilde{\mathcal{Z}}, \mathcal{I}, \boldsymbol{\lambda}) \leq \nu_F(\mathcal{Z}, \mathcal{I}, \boldsymbol{\lambda}) \leq \nu_F(\hat{\mathcal{Z}}, \mathcal{I}, \boldsymbol{\lambda}).$$

(ii) Usando a equação (3.61) da Proposição 21, deduz-se que

$$\begin{aligned} \nu_F(\tilde{\mathcal{Z}}, \mathcal{I}, \mathbf{1}) &= \frac{\varepsilon_{\tilde{\mathcal{Z}}}}{\varepsilon_{\tilde{\mathcal{Z}}} + 1} - \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} \frac{\varepsilon_{\tilde{\mathcal{Z}}_{\cup_{j \in S} I_j}}}{\varepsilon_{\tilde{\mathcal{Z}}_{\cup_{j \in S} I_j}} + 1} \\ &= \frac{1}{2} - \frac{1}{2} \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} \\ &\stackrel{(3.68)}{=} \frac{1}{2} - \frac{1}{2} \\ &= 0, \end{aligned}$$

já que \tilde{Z}_{I_j} , $j = 1, \dots, p$, são totalmente dependentes e, por isso, $\varepsilon_{\tilde{\mathcal{Z}}_{\cup_{j \in S} I_j}} = 1$, com $\emptyset \neq S \subseteq \{1, \dots, p\}$.

(iii) Recorrendo novamente à equação (3.61) da Proposição 21, obtém-se que

$$\begin{aligned} \nu_F(\hat{\mathcal{Z}}, \mathcal{I}, \mathbf{1}) &= \frac{\varepsilon_{\hat{\mathcal{Z}}}}{\varepsilon_{\hat{\mathcal{Z}}} + 1} - \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} \frac{\varepsilon_{\hat{\mathcal{Z}}_{\cup_{j \in S} I_j}}}{\varepsilon_{\hat{\mathcal{Z}}_{\cup_{j \in S} I_j}} + 1} \\ &= \frac{\sum_{j=1}^p \varepsilon_{\hat{Z}_{I_j}}}{\sum_{j=1}^p \varepsilon_{\hat{Z}_{I_j}} + 1} - \sum_{\emptyset \neq S \subseteq \{1, \dots, p\}} (-1)^{|S|+1} \frac{\sum_{j \in S} \varepsilon_{\hat{Z}_{I_j}}}{\sum_{j \in S} \varepsilon_{\hat{Z}_{I_j}} + 1}, \end{aligned}$$

uma vez que \hat{Z}_{I_j} são independentes para qualquer $j \in \{1, \dots, p\}$ e, por essa razão, $\varepsilon_{\hat{\mathcal{Z}}_{\cup_{j \in S} I_j}} = \sum_{j \in S} \varepsilon_{\hat{Z}_{I_j}}$, para $\emptyset \neq S \subseteq \{1, \dots, p\}$.

□

Observação 7. Seja $Z = \{Z(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$ um campo aleatório max-estável simples.

Considere-se $\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$, $\mathcal{I} = \{I_j = \{j\}, j \in \{1, \dots, p = k\}\}$ e $\lambda = \mathbf{1}$.

Se as variáveis do campo forem independentes, usando a alínea (iii) da proposição anterior, o madograma é dado por

$$\begin{aligned}
\nu_F(\mathbf{Z}, \mathcal{I}, \mathbf{1}) &= \frac{\sum_{j=1}^k \varepsilon_{Z_{\{j\}}}}{\sum_{j=1}^k \varepsilon_{Z_{\{j\}}} + 1} - \sum_{\emptyset \neq S \subseteq \{1, \dots, k\}} (-1)^{|S|+1} \frac{\sum_{j \in S} \varepsilon_{Z_{\{j\}}}}{\sum_{j \in S} \varepsilon_{Z_{\{j\}}} + 1} \\
&= \frac{k}{k+1} - \sum_{\emptyset \neq S \subseteq \{1, \dots, k\}} (-1)^{|S|+1} \frac{|S|}{|S|+1} \\
&= \frac{k}{k+1} - \sum_{j=1}^k (-1)^{j+1} \binom{k}{j} \frac{j}{j+1} \\
&= \frac{k}{k+1} + \frac{1}{k+1} \sum_{j=1}^k (-1)^{j+2} \binom{k}{j} \frac{j}{j+1} (k+1) \\
&= \frac{k}{k+1} + \frac{1}{k+1} \sum_{j=1}^k (-1)^{j+2} \frac{(k+1)!}{(j+1)!(k+1-(j+1))!} j \\
&= \frac{k}{k+1} + \frac{1}{k+1} \sum_{j=1}^k (-1)^{j+2} \binom{k+1}{j+1} j \\
&= \frac{k}{k+1} + \frac{1}{k+1} \sum_{j=1}^k (-1)^{j+2} \binom{k+1}{j+1} (j+1) - \frac{1}{k+1} \sum_{j=1}^k (-1)^{j+2} \binom{k+1}{j+1} \\
&= \frac{k}{k+1} + \frac{1}{k+1} \left(\sum_{j=0}^k (-1)^{j+2} \binom{k+1}{j+1} (j+1) - (-1)^2 \binom{k+1}{1} \right) \\
&\quad + \frac{1}{k+1} \sum_{j=1}^k (-1)^{j+1} \binom{k+1}{j+1} \\
&= \frac{k}{k+1} + \frac{1}{k+1} (-k-1) + \frac{1}{k+1} \left(\sum_{j=-1}^k (-1)^{j+1} \binom{k+1}{j+1} - (-1)^0 \binom{k+1}{0} - (-1)^1 \binom{k+1}{1} \right) \\
&= \frac{k}{k+1} - 1 + \frac{1}{k+1} (-1+k+1) \\
&= \frac{k-1}{k+1},
\end{aligned}$$

sendo que a antepenúltima igualdade se segue da identidade

$$\sum_{j=1}^{k+1} (-1)^{j+1} \binom{k+1}{j} j = 0,$$

enquanto a penúltima se deriva tendo em conta

$$\sum_{j=0}^{k+1} (-1)^j \binom{k+1}{j} = 0.$$

Relembrando as alíneas (i) e (ii) da Proposição 22, tem-se que

$$0 \leq \nu_F(\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)), \mathcal{I} = \{I_1 = \{1\}, \dots, I_k = \{k\}\}, \boldsymbol{\lambda} = \mathbf{1}) \leq \frac{k-1}{k+1},$$

onde os limites superior e inferior correspondem, respetivamente, aos casos em que as variáveis $Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)$ são totalmente dependentes e independentes.

Note-se que considerando a seguinte transformação

$$1 - \frac{k+1}{k-1} \nu_F(\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)), \mathcal{I} = \{I_1 = \{1\}, \dots, I_k = \{k\}\}, \boldsymbol{\lambda} = \mathbf{1}),$$

obtém-se o madograma de Ferreira e Pereira (2015) [18].

Estimação do madograma generalizado multiregional

Sejam $\mathbf{Z}^{(t)} = (Z^{(t)}(\mathbf{x}_1), \dots, Z^{(t)}(\mathbf{x}_k))$ uma amostra aleatória do vetor \mathbf{Z} e $F_{Z(\mathbf{x}_i)}$ a função de distribuição de $Z^{(t)}(\mathbf{x}_i)$, para $t \in \{1, \dots, T\}$ e $i \in \{1, \dots, k\}$. Aplicando o procedimento (3.24) de transformação das margens $Z^{(t)}(\mathbf{x}_i)$, tem-se que

$$\hat{\nu}_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda}) = \frac{1}{T} \sum_{t=1}^T \left[\bigvee_{j=1}^p \bigvee_{i \in I_j} F_{Z(\mathbf{x}_i)}^{\lambda_j}(Z^{(t)}(\mathbf{x}_i)) - \bigwedge_{j=1}^p \bigvee_{i \in I_j} F_{Z(\mathbf{x}_i)}^{\lambda_j}(Z^{(t)}(\mathbf{x}_i)) \right] \quad (3.74)$$

é um estimador fortemente consistente e assintoticamente normal para $\nu_F(\mathbf{Z}, \mathcal{I}, \boldsymbol{\lambda})$, cf. Ferreira (2013) [15], com $\mathcal{I} = \{I_1, I_2, \dots, I_p\}$ uma partição arbitrária de $\{1, \dots, k\}$, nas condições acima, e $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p) \in (0, +\infty)^p$.

Quando as funções de distribuição $F_{Z(\mathbf{x}_i)}$ de $Z^{(t)}(\mathbf{x}_i)$, para $t \in \{1, \dots, T\}$ e $i \in \{1, \dots, k\}$, são desconhecidas, substituem-se pelas correspondentes f.d. empíricas corrigidas.

Exemplo 19. Para ilustrar a aplicação prática do estimador anterior, considerámos três configurações distintas, cada uma composta por três regiões disjuntas de estações do conjunto de dados **rainfall**, definidas segundo diferentes critérios de agrupamento geográfico. O objetivo é investigar se os diferentes critérios de agrupamento geográfico traduzem diferentes níveis de homogeneidade no comportamento de precipitações elevadas e, assim, identificar que características geográficas são mais determinantes para a coocorrência de precipitações elevadas e, conseqüentemente, mais informativas na definição de agrupamentos homogêneos.

A escolha de limitar cada configuração a apenas três regiões visa evitar a complexidade combinatória inerente a um maior número de regiões e assegurar uma leitura mais clara

dos resultados obtidos.

À semelhança do que foi feito nos dois madogramas anteriores, esta análise baseia-se em subconjuntos restritos das 79 estações disponíveis. As conclusões devem, portanto, ser vistas como ilustrativas e condicionadas pela seleção efetuada.

As três configurações consideradas são as seguintes:

- *Configuração A: Agrupamento por corpos de água, com $I_1 = \{1, 2\}$, sendo $x_1 = l_{34}$ e $x_2 = l_{39}$; $I_2 = \{3\}$, para $x_3 = l_{77}$; e $I_3 = \{4\}$, com $x_4 = l_{70}$.*

- *Configuração B: Agrupamento numa planície, com $I_1 = \{1\}$, para $x_1 = l_5$; $I_2 = \{2, 3\}$, com $x_2 = l_{13}$ e $x_3 = l_{19}$; e $I_3 = \{4\}$, sendo $x_4 = l_{15}$.*

- *Configuração C: Agrupamento por regiões montanhosas, com $I_1 = \{1, 2, 3\}$, tendo $x_1 = l_{49}$, $x_2 = l_{20}$ e $x_3 = l_{30}$; $I_2 = \{4, 5\}$, sendo $x_4 = l_{75}$ e $x_5 = l_{53}$; e $I_3 = \{6\}$, com $x_6 = l_{61}$.*

A tabela 3.4 apresenta as coordenadas e altitudes das estações com as quais ainda não trabalhamos.

Tabela 3.4: Longitudes, latitudes e altitudes das estações 34, 39, 61, 70 e 77.

Localização	Longitude (m)	Latitude (m)	Altitude (m)
l_{34}	676.325	215.325	450
l_{39}	665.540	209.848	454
l_{61}	737.700	227.600	890
l_{70}	725.245	221.680	435
l_{77}	693.849	230.708	485

As estimativas obtidas para as configurações A, B e C são apresentadas nas Figuras 3.17, 3.18 e 3.19, respetivamente. Analisemo-las:

Considerando pesos iguais para as regiões, ou seja, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda \in (0, 100)$,

(i) é observável que, para qualquer configuração, as estimativas se aproximam de 0 à medida que λ aumenta, o que evidencia que valores maiores de λ reduzem a sensibilidade à ocorrência simultânea de níveis elevados de precipitação nas regiões;

(ii) comparando as configurações: a configuração A apresenta estimativas mais elevadas, sugerindo menor dependência entre as precipitações elevadas nas regiões nela consideradas. A configuração B regista os valores mais baixos, refletindo maior homogeneidade dos níveis de precipitação elevada das suas regiões. A configuração C surge numa posição intermédia.

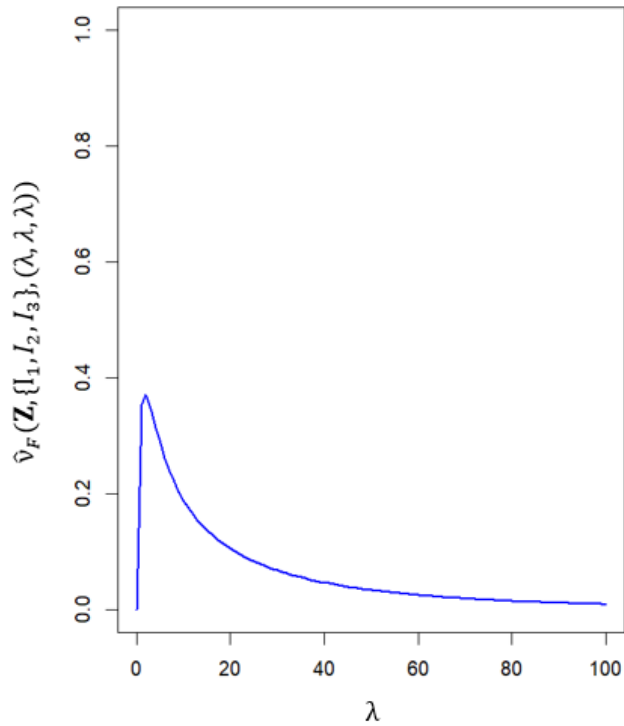


Figura 3.17: Estimativas de $\hat{\nu}_F(\mathbf{Z}, \{I_1, I_2, I_3\}, (\lambda, \lambda, \lambda))$ para configuração A do Exemplo 19 com $\lambda \in (0, 100)$.

Concluimos, neste enquadramento, que diferentes critérios de agrupamento geográfico refletem, de facto, diferentes estruturas de dependência entre precipitações elevadas. Por um lado, as regiões agrupadas por bacias hidrográficas distintas mostram menor homogeneidade, sugerindo que a pertença a corpos de água diferentes pode ser um fator de desacoplamento na ocorrência simultânea de precipitações elevadas. Por outro lado, as configurações que agregam regiões inseridas em planícies tendem a apresentar maior homogeneidade, refletida em menores valores do estimador, indicando uma maior propensão à coocorrência de precipitações elevadas nestas áreas.

Este tipo de análise é particularmente relevante na definição de políticas de gestão de risco, pois permite identificar áreas com maior vulnerabilidade conjunta a acontecimentos extremos e, assim, orientar medidas preventivas e recursos para as regiões mais críticas.

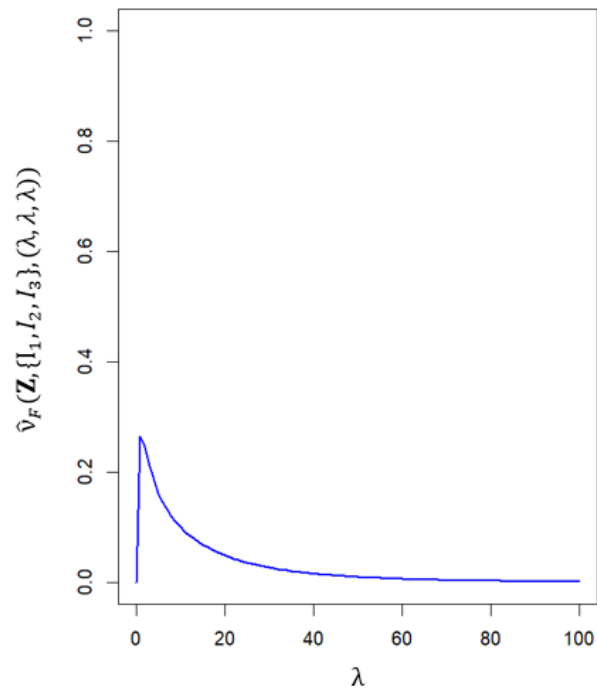


Figura 3.18: Estimativas de $\hat{\nu}_F(\mathbf{Z}, \{I_1, I_2, I_3\}, (\lambda, \lambda, \lambda))$ para configuração B do Exemplo 19 com $\lambda \in (0, 100)$.

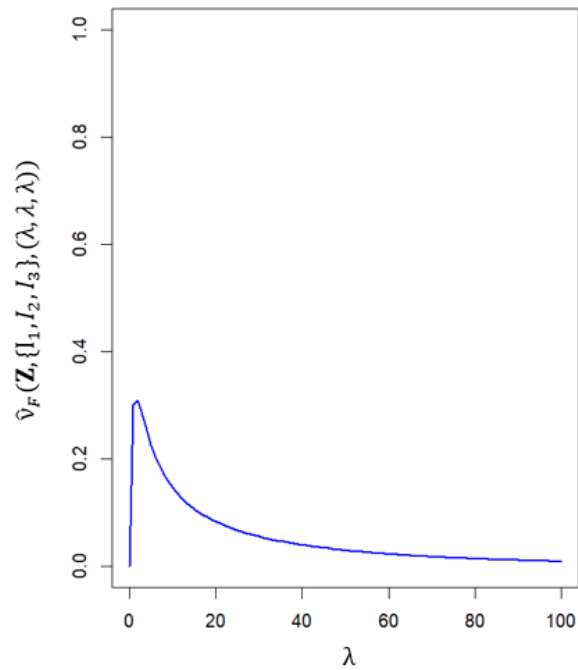


Figura 3.19: Estimativas de $\hat{\nu}_F(\mathbf{Z}, \{I_1, I_2, I_3\}, (\lambda, \lambda, \lambda))$ para configuração C do Exemplo 19 com $\lambda \in (0, 100)$.

Conclusões

Ao longo desta dissertação procurámos caracterizar a dependência entre acontecimentos extremos num contexto espacial recorrendo, por um lado, aos campos max-estáveis para modelar o contexto espacial e, por outro, aos madogramas para caracterizar a dependência nesse contexto. Os madogramas e as suas relações com medidas de dependência da Teoria de Valores Extremos revelaram-se particularmente úteis para esse propósito. Contudo, a classe de campos aleatórios max-estáveis tem uma limitação: captura apenas a dependência assintótica, excluindo cenários de independência assintótica. Nesse contexto, a classe de campos max-estáveis invertidos, propostos por Wadsworth e Tawn (2012) [46], surge como uma alternativa metodologicamente mais adequada. Não obstante, existe ainda um terceiro cenário relevante: a coexistência da dependência e independência assintótica. Para este último caso, emerge a classe de campos híbridos, também propostos por Wadsworth e Tawn (2012) [46], que, pela sua flexibilidade, representa uma abordagem promissora para investigação futura. A capacidade destes modelos representarem, simultaneamente, dependência e independência assintótica, amplia significativamente o leque de aplicações na modelação de acontecimentos extremos em contextos espaciais. Além disso, a integração e adaptação dos madogramas para estes modelos possibilitaria a criação de medidas de dependência mais robustas e adaptadas à complexidade dos fenómenos observados.

Bibliografia

- [1] Brown, B. e Resnick, S. (1977). Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14(4), 732–739.
- [2] Buishand, T. A., de Haan, L. e Zhou, C. (2008). On spatial extremes: with application to rainfall problem. *Annals of Applied Statistics*, 2(2), 624–642.
- [3] Carvalho, M. Lucília e Natário, Isabel C. (2008). *Análise de Dados Espaciais*, Sociedade Portuguesa de Estatística.
- [4] Coles, S. G. e Tawn, J. A. (1996). Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society: Series B*, 58(2), 329–347.
- [5] Cooley, D., Naveau, P. e Poncet, P. (2006). Variograms for spatial max-stable random fields. In: Bertail, P., Soulier, P., Doukhan, P. (eds) *Dependence in Probability and Statistics. Lecture Notes in Statistics*, 187, 373–390. Springer, New York.
- [6] Cressie, Noel A. C. (1993). *Statistics for Spatial Data*, John Wiley & Sons, New York.
- [7] de Haan, L. (1970). *On regular variation and its applications to the weak convergence of sample extremes* [Tese de Doutorado, Universidade de Amesterdão]. Mathematisch Centrum. Mathematical Centre Tracts, 32.
- [8] de Haan, L. (1984). A Spectral Representation for Max-stable Processes. *The Annals of Probability*, 12(4), 1194–1204.
- [9] de Haan, L. e Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering.
- [10] de Haan, L. e Resnick, S. (1977). Limit theory for multivariate sample extremes. *Z. Wahrsch. verw. Gebiete*, 40, 317–337.
- [11] Deheuvels, P. (1978). Caractérisation complète des lois extrêmes multivariées de la convergence aux types extrêmes. *Publications de l'Institut de statistique de l'Université de Paris*, 23, 1–36.
- [12] Deheuvels, P. (1984). *Probabilistic aspects of multivariate extremes*. In de Oliveira, J.T. (Ed.) *Statistical Extremes and Applications* (pp. 117–130). Reidel.
- [13] Esary, J. D., Proschan, F. e Walkup, D. W. (1967). Association of Random Variables, with Applications. *The Annals of Mathematical Statistics*, 38(1), 1466–1474.

- [14] Ferreira, H. (2011). Dependence between two multivariate extremes. *Statistics & Probability Letters*, 81(5), 586–591.
- [15] Ferreira, H. (2013) Max-min dependence coefficients for Multivariate Extreme Value Distributions. arXiv:1205.3954v2.
- [16] Ferreira, H. e Ferreira, M. (2012). On extremal dependence of block vectors. *Kybernetika*, 48(5), 988–1006.
- [17] Ferreira, H. e Ferreira, M. (2018). Multidimensional extremal dependence coefficients. *Statistics & Probability Letters*, 133, 1–8.
- [18] Ferreira, H. e Pereira, L. (2015). Dependence of maxima in space. *Journal of Physics: Conference Series*, 574, 012021.
- [19] Fonseca, C. (2012). *Coefficientes de dependência em campos aleatórios max-estáveis* [Tese de Doutorado, Universidade da Beira Interior].
- [20] Fonseca, C., Martins, A. P., Pereira, L. e Ferreira, H. (2016). Dependence matrices for spatial extreme events. *Communications in Statistics – Theory and Methods*, 45(21), 6321–6341.
- [21] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics* (2nd ed.), Robert E. Krieger Publishing Co. Inc., Melbourne, FL.
- [22] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics, Second Series*, 44(3), 423–453.
- [23] Greenwood, J. A., Landwehr, J. M., Matalas, N. C. e Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15(5), 1049–1054.
- [24] Hosking, J. R. M., Wallis, J. R. e Wood. E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3), 251–261.
- [25] Hsing, T. (1989). Extreme value theory for multivariate stationary sequences. *Journal of Multivariate Analysis*, 29 (2), 274–291.
- [26] Kabluchko, Z. (2009). Extremes of space-time Gaussian processes. *Stochastic Processes and their Applications*, 119(11), 3962–3980.
- [27] Leadbetter, M. R., Lindgren, G. e Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag New York.
- [28] Lehmann, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, 37(5), 1137–1153.
- [29] Lelowitz, J. (1972). Bounds on the correlations and analyticity of ferromagnetic Ising spin systems. *Communications in Mathematical Physics*, 28, 313–321.

- [30] Li, H. (2009). Orthant tail dependence of multivariate extreme value distributions. *Journal of Multivariate Analysis*, 100(1), 243–256.
- [31] Marcon, G., Padoan, S. A., Naveau, P., Muliere, P. e Segers, P. (2017) Multivariate non-parametric estimation of the Pickands dependence function using Bernstein polynomials. *Journal of Statistical Planning and Inference*, 183, 1–17.
- [32] Marshall, A. W. e Olkin, I. (1983). Domains of Attraction of Multivariate Extreme Value Distributions. *The Annals of Probability*, 11(1), 168–177.
- [33] Matheron, G. (1987). Suffit-il, pour une covariance, d’être de type positif? *Sciences de la Terre, série informatique géologique*, 26, 51–66.
- [34] Naveau, P., Guillou, A., Cooley, D. e Diebolt, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika*, 96(1), 1–17.
- [35] Oliveira, J. T. (1962-1963). Structure theory of bivariate extremes: extensions. *Estudos de Matemática, Estatística e Econometria*, 7, 165–195.
- [36] Padoan, S. A., Ribatet M. e Sisson, S. A. (2010). Likelihood-Based Inference for Max-Stable Processes. *Journal of the American Statistical Association*, 105(489), 263–277.
- [37] Pickands, J. (1981). Multivariate extreme value distributions. *Bulletin of the International Statistical Institute*, 49, 859–878.
- [38] Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York.
- [39] Schlather, M. (2002). Models for Stationary Max-Stable Random Fields. *Extremes*, 5, 33–44.
- [40] Schlather, M. e Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and Inference. *Biometrika*, 90(1), 139–156.
- [41] Shaked, M. e Shanthikumar, J.G. (2007). *Stochastic orders*. Springer-Verlag, New York.
- [42] Simpson, E. S., Wadsworth, J. L. e Tawn, J. A. (2020). Determining the dependence structure of multivariate extremes. *Biometrika*, 107(3), 513–532.
- [43] Sklar, M. (1959). Fonctions de répartition a N dimensions et leurs marges. *Annales de l’ISUP*, 8(3), 229–231.
- [44] Smith, R. L. (1990). *Max-stable processes and spatial extremes*. University of Surrey.
- [45] Smith, R. L. e Weissman, I. (1996). *Characterization and estimation of the multivariate extremal index*. Technical Report, University of North Carolina.
- [46] Wadsworth, J. L. e Tawn, J. A. (2012). Dependence modelling for spatial extremes. *Biometrika*, 99(2), 253–272.

Apêndice

Proposição 23. *Nas condições da Proposição 10,*

$$P(Z(\mathbf{x}_1) \leq z_1, \dots, Z(\mathbf{x}_k) \leq z_k) = \exp \left(-E \left[\bigvee_{j=1}^k \frac{Y(\mathbf{x}_j)}{z_j} \right] \right),$$

para $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{D}$, $z_1, \dots, z_k \in \mathbb{R}^+$ e $k \in \mathbb{N}$.

Demonstração. Para $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{D}$, $z_1, \dots, z_k \in \mathbb{R}^+$ e $k \in \mathbb{N}$,

$$\begin{aligned} P(Z(\mathbf{x}_1) \leq z_1, \dots, Z(\mathbf{x}_k) \leq z_k) &= P \left(\bigvee_{i \in \mathbb{N}} V_i Y_i(\mathbf{x}_1) \leq z_1, \dots, \bigvee_{i \in \mathbb{N}} V_i Y_i(\mathbf{x}_k) \leq z_k \right) \\ &= P(V_i Y_i(\mathbf{x}_1) \leq z_1, \dots, V_i Y_i(\mathbf{x}_k) \leq z_k, \forall i \in \mathbb{N}) \\ &= P \left(V_i \leq \frac{z_1}{Y_i(\mathbf{x}_1)}, \dots, V_i \leq \frac{z_k}{Y_i(\mathbf{x}_k)}, \forall i \in \mathbb{N} \right) \\ &= P \left(V_i \leq \bigwedge_{j=1}^k \frac{z_j}{Y_i(\mathbf{x}_j)}, \forall i \in \mathbb{N} \right) \\ &= P \left(N_V \left(\bigwedge_{j=1}^k \frac{z_j}{Y(\mathbf{x}_j)}, +\infty \right) = 0 \right) \\ &= \exp \left(- \int \int_{\bigwedge_{j=1}^k \frac{z_j}{Y(\mathbf{x}_j)}}^{+\infty} \frac{d\nu}{\nu^2} dF_Y \right) \\ &= \exp \left(- \int \left[-\frac{1}{\nu} \right]_{\bigwedge_{j=1}^k \frac{z_j}{Y(\mathbf{x}_j)}}^{+\infty} dF_Y \right) \\ &= \exp \left(- \int \frac{1}{\bigwedge_{j=1}^k \frac{z_j}{Y(\mathbf{x}_j)}} dF_Y \right) \\ &= \exp \left(-E_Y \left[\frac{1}{\bigwedge_{j=1}^k \frac{z_j}{Y(\mathbf{x}_j)}} \right] \right) \\ &= \exp \left(-E_Y \left[\bigvee_{j=1}^k \frac{Y(\mathbf{x}_j)}{z_j} \right] \right), \end{aligned}$$

tendo em conta que $N_V(\cdot) := \sum_{i \in \mathbb{N}} \delta_{V_i}(\cdot)$ é um processo pontual de Poisson definido em $(0, +\infty)$, de intensidade $\nu^{-2} d\nu$, e $Y_i, i \in \mathbb{N}$, são cópias independentes do campo aleatório $Y = \{Y(\mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}$. F_Y representa a f.d. de $Y(\mathbf{x}_j)$, com $\mathbf{x}_j \in \mathcal{D}$, para $j \in \{1, \dots, k\}$. □

Cálculos auxiliares ao Exemplo 18

Dado o campo e a partição apresentados, tem-se que

$$\begin{aligned}
\bullet \sum_{m=1}^2 \bigvee_{i \in I_1} \lambda_1^{-1} a_{1,m,\mathbf{x}_i} &= \bigvee_{i \in I_1} \lambda_1^{-1} a_{1,1,\mathbf{x}_i} + \bigvee_{i \in I_1} \lambda_1^{-1} a_{1,2,\mathbf{x}_i} \\
&= \lambda_1^{-1} (a_{1,1,(2,1)} \vee a_{1,1,(3,1)} \vee a_{1,1,(4,1)} \vee a_{1,1,(3,2)} \vee a_{1,1,(4,2)} \vee a_{1,1,(4,3)}) \\
&\quad + \lambda_1^{-1} (a_{1,2,(2,1)} \vee a_{1,2,(3,1)} \vee a_{1,2,(4,1)} \vee a_{1,2,(3,2)} \vee a_{1,2,(4,2)} \vee a_{1,2,(4,3)}) \\
&= \lambda_1^{-1} \left(\frac{1}{3} \vee \frac{2}{3} \vee \frac{1}{3} \vee \frac{1}{3} \vee \frac{1}{2} \vee \frac{1}{3} \right) \\
&\quad + \lambda_1^{-1} \left(\frac{2}{3} \vee \frac{1}{3} \vee \frac{2}{3} \vee \frac{2}{3} \vee \frac{1}{2} \vee \frac{2}{3} \right) \\
&= \frac{2}{3} \lambda_1^{-1} + \frac{2}{3} \lambda_1^{-1} \\
&= \frac{4}{3} \lambda_1^{-1},
\end{aligned}$$

$$\begin{aligned}
\bullet \sum_{m=1}^2 \bigvee_{i \in I_2} \lambda_2^{-1} a_{1,m,\mathbf{x}_i} &= \bigvee_{i \in I_2} \lambda_2^{-1} a_{1,1,\mathbf{x}_i} + \bigvee_{i \in I_2} \lambda_2^{-1} a_{1,2,\mathbf{x}_i} \\
&= \lambda_2^{-1} (a_{1,1,(1,1)} \vee a_{1,1,(2,2)} \vee a_{1,1,(3,3)} \vee a_{1,1,(4,4)}) \\
&\quad + \lambda_2^{-1} (a_{1,2,(1,1)} \vee a_{1,2,(2,2)} \vee a_{1,2,(3,3)} \vee a_{1,2,(4,4)}) \\
&= \lambda_2^{-1} \left(\frac{2}{3} \vee \frac{1}{2} \vee \frac{2}{3} \vee \frac{1}{2} \right) \\
&\quad + \lambda_2^{-1} \left(\frac{1}{3} \vee \frac{1}{2} \vee \frac{1}{3} \vee \frac{1}{2} \right) \\
&= \frac{2}{3} \lambda_2^{-1} + \frac{1}{2} \lambda_2^{-1} \\
&= \frac{7}{6} \lambda_2^{-1},
\end{aligned}$$

e

$$\begin{aligned}
\bullet \sum_{m=1}^2 \bigvee_{i \in I_3} \lambda_3^{-1} a_{1,m,\mathbf{x}_i} &= \bigvee_{i \in I_3} \lambda_3^{-1} a_{1,1,\mathbf{x}_i} + \bigvee_{i \in I_3} \lambda_3^{-1} a_{1,2,\mathbf{x}_i} \\
&= \lambda_3^{-1} (a_{1,1,(1,2)} \vee a_{1,1,(1,3)} \vee a_{1,1,(2,3)} \vee a_{1,1,(1,4)} \vee a_{1,1,(2,4)} \vee a_{1,1,(3,4)}) \\
&\quad + \lambda_3^{-1} (a_{1,2,(1,2)} \vee a_{1,2,(1,3)} \vee a_{1,2,(2,3)} \vee a_{1,2,(1,4)} \vee a_{1,2,(2,4)} \vee a_{1,2,(3,4)}) \\
&= \lambda_3^{-1} \left(\frac{1}{3} \vee \frac{2}{3} \vee \frac{1}{3} \vee \frac{1}{3} \vee \frac{1}{2} \vee \frac{1}{3} \right) \\
&\quad + \lambda_3^{-1} \left(\frac{2}{3} \vee \frac{1}{3} \vee \frac{2}{3} \vee \frac{2}{3} \vee \frac{1}{2} \vee \frac{2}{3} \right) \\
&= \frac{2}{3} \lambda_3^{-1} + \frac{2}{3} \lambda_3^{-1} \\
&= \frac{4}{3} \lambda_3^{-1};
\end{aligned}$$

Donde,

$$\begin{aligned}
\bullet \sum_{m=1}^2 \bigvee_{j \in \{1,2\}} \bigvee_{i \in I_j} \lambda_j^{-1} a_{1,m,\mathbf{x}_i} &= \bigvee_{j \in \{1,2\}} \bigvee_{i \in I_j} \lambda_j^{-1} a_{1,1,\mathbf{x}_i} + \bigvee_{j \in \{1,2\}} \bigvee_{i \in I_j} \lambda_j^{-1} a_{1,2,\mathbf{x}_i} \\
&= \bigvee_{i \in I_1} \lambda_1^{-1} a_{1,1,\mathbf{x}_i} \vee \bigvee_{i \in I_2} \lambda_2^{-1} a_{1,1,\mathbf{x}_i} + \bigvee_{i \in I_1} \lambda_1^{-1} a_{1,2,\mathbf{x}_i} \vee \bigvee_{i \in I_2} \lambda_2^{-1} a_{1,2,\mathbf{x}_i} \\
&= \frac{2}{3} \lambda_1^{-1} \vee \frac{2}{3} \lambda_2^{-1} + \frac{2}{3} \lambda_1^{-1} \vee \frac{1}{2} \lambda_2^{-1} \\
&= \frac{2}{3} (\lambda_1^{-1} \vee \lambda_2^{-1}) + \frac{2}{3} \lambda_1^{-1} \vee \frac{1}{2} \lambda_2^{-1},
\end{aligned}$$

$$\begin{aligned}
\bullet \sum_{m=1}^2 \bigvee_{j \in \{1,3\}} \bigvee_{i \in I_j} \lambda_j^{-1} a_{1,m,\mathbf{x}_i} &= \bigvee_{j \in \{1,3\}} \bigvee_{i \in I_j} \lambda_j^{-1} a_{1,1,\mathbf{x}_i} + \bigvee_{j \in \{1,3\}} \bigvee_{i \in I_j} \lambda_j^{-1} a_{1,2,\mathbf{x}_i} \\
&= \bigvee_{i \in I_1} \lambda_1^{-1} a_{1,1,\mathbf{x}_i} \vee \bigvee_{i \in I_3} \lambda_3^{-1} a_{1,1,\mathbf{x}_i} + \bigvee_{i \in I_1} \lambda_1^{-1} a_{1,2,\mathbf{x}_i} \vee \bigvee_{i \in I_3} \lambda_3^{-1} a_{1,2,\mathbf{x}_i} \\
&= \frac{2}{3} \lambda_1^{-1} \vee \frac{2}{3} \lambda_3^{-1} + \frac{2}{3} \lambda_1^{-1} \vee \frac{2}{3} \lambda_3^{-1} \\
&= \frac{2}{3} (\lambda_1^{-1} \vee \lambda_3^{-1}) + \frac{2}{3} (\lambda_1^{-1} \vee \lambda_3^{-1}) \\
&= \frac{4}{3} (\lambda_1^{-1} \vee \lambda_3^{-1}),
\end{aligned}$$

e

$$\begin{aligned}
\bullet \sum_{m=1}^2 \bigvee_{j \in \{2,3\}} \bigvee_{i \in I_j} \lambda_j^{-1} a_{1,m,\mathbf{x}_i} &= \bigvee_{j \in \{2,3\}} \bigvee_{i \in I_j} \lambda_j^{-1} a_{1,1,\mathbf{x}_i} + \bigvee_{j \in \{2,3\}} \bigvee_{i \in I_j} \lambda_j^{-1} a_{1,2,\mathbf{x}_i} \\
&= \bigvee_{i \in I_2} \lambda_2^{-1} a_{1,1,\mathbf{x}_i} \vee \bigvee_{i \in I_3} \lambda_3^{-1} a_{1,1,\mathbf{x}_i} + \bigvee_{i \in I_2} \lambda_2^{-1} a_{1,2,\mathbf{x}_i} \vee \bigvee_{i \in I_3} \lambda_3^{-1} a_{1,2,\mathbf{x}_i} \\
&= \frac{2}{3} \lambda_2^{-1} \vee \frac{2}{3} \lambda_3^{-1} + \frac{1}{2} \lambda_2^{-1} \vee \frac{2}{3} \lambda_3^{-1} \\
&= \frac{2}{3} (\lambda_2^{-1} \vee \lambda_3^{-1}) + \frac{1}{2} \lambda_2^{-1} \vee \frac{2}{3} \lambda_3^{-1}.
\end{aligned}$$

