CONGRESSO ANUAL

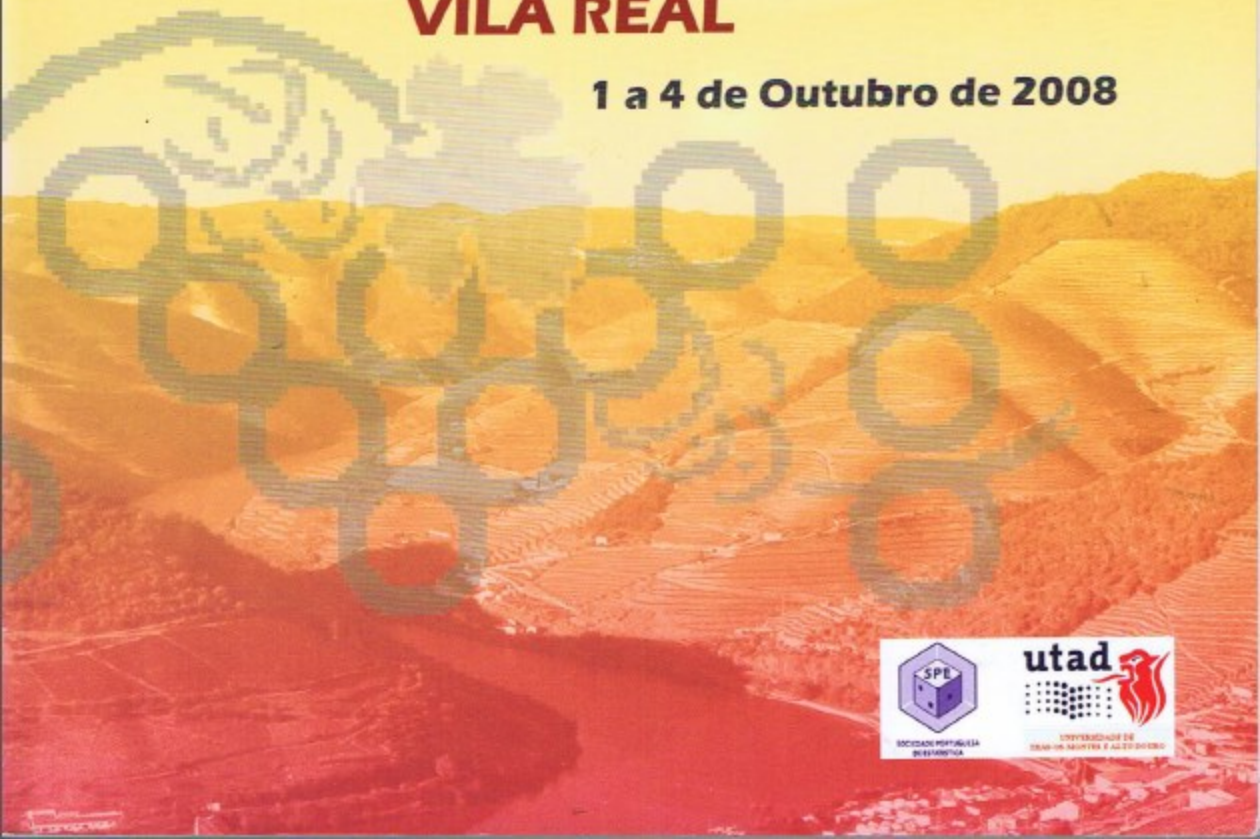
Sociedade Portuguesa de Estatística

Análise de Dados Espaciais

M. Lucília Carvalho e Isabel C. Natário

VILA REAL

1 a 4 de Outubro de 2008



Prefácio

Este livro foi escrito para servir de base ao mini-curso do XVI Congresso da Sociedade Portuguesa de Estatística sobre dados espaciais.

O reconhecimento de uma natureza espacial em dados como a distribuição de minérios no subsolo, a concentração de poluentes na atmosfera, a localização de focos de disseminação de doenças, o número de casos de doenças cuja etiologia depende da localização da morada dos doentes, etc., de uma forma que não apenas a da construção de mapas que os representem, levou a que a comunidade científica tivesse de formular modelos e teorias que conseguissem descrever, modelar e prever estes fenómenos dependentes da sua localização no espaço.

Espacial é especial!

Estes dados classificam-se tipicamente em três categorias, conforme o problema em questão. Por vezes estamos interessados num fenómeno que ocorre em toda uma região no espaço, mas que apenas é amostrado num subconjunto de localizações - dados por pontos (ou dados geoestatísticos). Outras vezes usamos dados que foram recolhidos de forma agregada em subdivisões da região em estudo - dados agregados. Outras vezes ainda a localização espacial do fenómeno de

interesse é o que se pretende modelar - padrões pontuais.

O objectivo deste curso é fazer uma introdução aos dados espaciais e às principais técnicas e modelos que existem para os tratar. Apesar de remontar apenas à segunda metade do século XX a aceitação da necessidade de encontrar formas específicas para efectuar este tratamento, a verdade é que é já vasta a literatura sobre o assunto. Não é nosso propósito descrevê-la exaustivamente. Pretendemos sim apresentar as principais linhas de acção e investigação da estatística espacial, tentando dar uma compreensão global destes conceitos e métodos, ilustrando-as com alguns exemplos práticos.

M. Lucília Carvalho e Isabel C. Natário

Lisboa, 31 de Julho de 2008

Agradecimentos

Agradecemos à Sociedade Portuguesa de Estatística a oportunidade que nos deu de publicar estas notas e de ministrar o mini-curso sobre um assunto que nos é tão grato.

M. L. C. e I. C. N.

Índice

1 Introdução **1**

1.1 Um pouco de história 1

1.2 Objectivos da análise de dados espaciais 6

1.3 Modelos e realidade 13

1.3.1 Cálculo de distâncias sobre o globo terrestre 14

2 Modelos referentes a pontos **21**

2.1 Nomenclatura e conceitos básicos 21

2.1.1 Hipóteses iniciais 21

2.1.2 Estacionaridade dos processos espaciais 22

2.1.3 Propriedades do variograma e covariograma 26

2.1.4 Isotropia e Anisotropia 33

2.1.5 Correlograma e diferenciabilidade do processo 37

2.1.6 Famílias paramétricas de variogramas 40

2.2 Elementos da modelação 48

2.2.1 Modelação da tendência espacial 50

2.2.2 Variograma empírico 52

2.3 Estimação 60

2.3.1 Estimação da tendência espacial 61

2.3.2 Estimação do variograma 63

2.4 Predição espacial clássica 79

2.4.1 Considerações finais 88

3 Modelos Referentes a Áreas 91

3.1 Questões básicas sobre dados referentes a áreas 91

3.2 Padrão espacial 95

3.2.1 Medidas de associação espacial 98

3.3 Alisamento espacial 107

3.4 Campos Aleatórios de Markov 113

3.4.1 Cliques, Potênciais e Distribuições de Gibbs 118

3.5 Modelação de processos referentes a áreas 121

3.5.1 Modelos auto-regressivos condicionais (CAR) 122

3.5.2 Modelos auto-regressivos simultâneos (SAR) 130

3.5.3 Comparação dos modelos CAR e SAR 133

3.5.4 Aplicações 134

4 Processos Pontuais Espaciais 139

4.1 Introdução e exemplos 139

4.2 Processos pontuais espaciais 144

4.2.1 Nomenclatura e conceitos básicos 144

4.2.2 Processos pontuais marcados 149

4.2.3 Propriedades de segunda ordem 150

4.2.4 Propriedades de ordem superior e distribuições de vizinhos mais próximos 155
4.2.5 O Processo de Poisson homogéneo 155
4.2.6 O Processo de Poisson não homogéneo 159
4.2.7 Aplicações 161
4.2.8 Estimação das propriedades de segunda ordem 163
4.2.9 Aplicações 168
4.2.10 Estimação das funções distribuição do espaço vazio e do vizinho mais próximo 171
4.2.11 Aplicações 173
4.3 Teste à hipótese de aleatoriedade espacial completa 174
4.4 Modelos 181
4.4.1 Processos de Poisson agregados 182
4.4.2 Processos de Cox 184
4.4.3 Processos de inibição simples 186
4.4.4 Processos pontuais de Markov 188
4.5 Ajustamento de modelos 192
4.5.1 Ajustamento de modelos usando medidas descritivas 192
4.5.2 Ajustamento de modelos baseada na máxima verosimilhança 195
4.5.3 Aplicações 201
4.6 Processos pontuais espaço-temporais 202

Referências **211**

Capítulo 1

Introdução

1.1 Um pouco de história

Um exemplo de como um método adequado de representar dados, aqui eminentemente espacial, pode indicar a solução de um problema complexo, mesmo sem se conhecer exactamente as causas do fenómeno em análise, está bem patente no trabalho de John Snow (1855) sobre a epidemia de cólera, que alastrou em Londres no século XIX, e que descrevemos brevemente.

A 31 de Agosto de 1854 surgiu no centro de Londres uma epidemia de cólera que matou mais de 500 pessoas em cerca de 10 dias. O alarme foi geral e os londrinos abandonaram a cidade em massa o que provocou uma forte interrupção da vida económica e social da cidade.

O diagrama da Figura 1.1, onde se representa o número de mortes por cólera registadas diariamente durante a epidemia, mostra bem o rápido desenvolvimento da doença.

O panorama era tanto mais assustador quanto à data nada se sabia sobre a causa da cólera e se pensava que a doença se espalhava através do ar ou, mais fantasiosamente, que subia em forma de vapor dos locais onde tinham sido enterradas vítimas da doença 2 séculos antes.

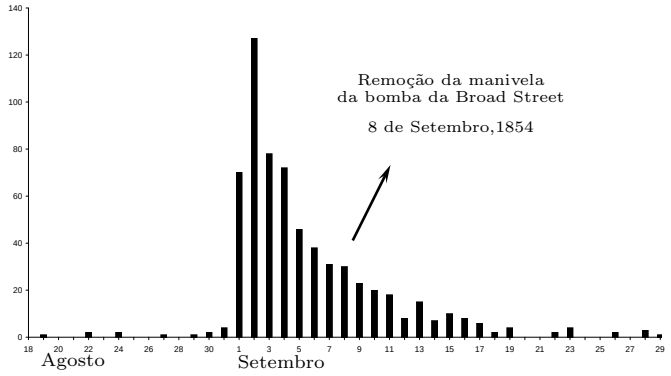


Figura 1.1 *Mortes por cólera em cada dia da epidemia*

No entanto, John Snow estava convencido que a doença era transmitida através de água contaminada e, para tentar provar essa teoria, registou no **mapa** da parte de Londres onde a epidemia alastrava, a **localização** precisa da morada de cada um dos 83 mortos por cólera de uma lista que obteve do General Register Office.

O mapa que desenhou, do qual apresentamos na Figura 1.2 uma cópia, mostra claramente que a maioria das mortes, assinaladas com pequenos traços negros nas respectivas residências, se verificaram nas imediações do poço situado na confluência da Broad Street com a Cambridge Street, marcado com dois círculos no mapa.

Apesar do mapa ser bastante informativo, John Snow fez uma investigação porta a porta sobre os hábitos de abastecimento de água dos 83 dos casos, onde pôde confirmar a correlação existente entre as mortes e a utilização de água do poço da Broad Street.

Embora tivesse observado que a água do poço não apresentava impurezas suspeitas, Snow descreveu as suas descobertas às autoridades responsáveis pelo abastecimento de água a 7 de Setembro e, no dia

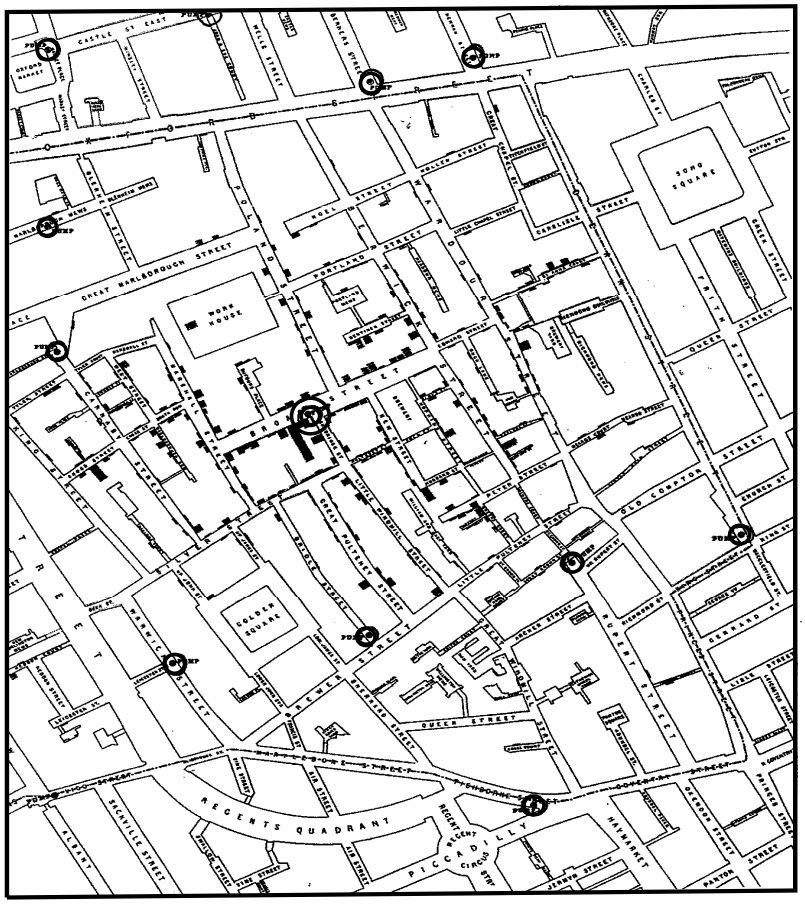


Figura 1.2 *Mapa de Londres onde deflagrou a epidemia:*
1 círculo - poço não infectado;
2 círculos - poço infectado da Broad Street;
■ *- nº de mortes por cólera no local, proporcional ao comprimento do rectângulo*

seguinte, a manivela da bomba do referido poço foi removida.

É erradamente atribuída à remoção do manípulo o fim deste surto

de cólera. De facto, como se pode verificar na Figura 1.1, a epidemia nessa data já estava na sua fase descendente, muito provavelmente devido à fuga do local de grande parte da população. No entanto, a não utilização do poço contaminado foi certamente responsável pela prevenção de novos surtos quando os ocupantes regressaram às suas residências.

Em 1886 a descoberta do vibrião da cólera veio confirmar a hipótese tão brilhantemente avançada por Snow, cujo eloquente trabalho é com frequência apontado como o primeiro exemplo de aplicação da representação espacial de dados na descoberta de solução para um problema de saúde pública.

Mais recentemente, outro problema prático deu origem ao desenvolvimento de outro ramo de análise espacial de dados que é habitualmente conhecido por **Geoestatística**.

Trata-se da avaliação da rentabilidade de exploração de uma mina através da estimação da quantidade de minério aí existente feita com base em amostras retiradas de locais igualmente espaçados no terreno, usando um método empírico de cálculo apresentado em 1951 pelo engenheiro de minas sul-africano D. J. Krige.

Este problema da predição espacial, foi mais tarde tratado do ponto de vista mais formal por G. Matheron (1962) em França, sempre no contexto da exploração mineira, e por Gandin (1963) na União Soviética, no contexto da meteorologia, e desde então não tem deixado de conhecer enormes desenvolvimentos com a utilização, cada vez mais vulgar, de sofisticados meios computacionais e de representação geográfica fornecidos pelos Sistemas de Informação Geográfica (SIG).

Outro tipo de problema de dependência espacial foi certamente perce-

bido por R. A. Fisher durante o seu trabalho de agricultura experimental, nas décadas de 20 e 30 do século XX na Rothamsted Experimental Station em Inglaterra, que deu origem ao planeamento de experiências. De facto, a par dos princípios de divisão em blocos e de replicação, a aleatorização foi criada para controlar o enviesamento e também para neutralizar (embora não remova) o efeito da correlação espacial.

Este tipo de **dados que se referem a áreas** contíguas de formato regular ou irregular foram tratados extensivamente por Besag (1974, 1975) e os modelos que lhes estão associados fundamentam-se não na noção de distância, como nos casos descritos anteriormente, mas sim na noção de vizinhança entre regiões. Estes modelos experimentaram desde então um enorme desenvolvimento visto serem muito adequados ao estudo e tratamento de imagens, quer das fornecidas por satélite na detecção remota, quer das fornecidas por aparelhos de imagem para diagnóstico médico, mas neste curso estamos especialmente interessados na sua aplicação no campo da epidemiologia espacial.

Os 3 tipos de dados espaciais que descrevemos serão objecto de um maior detalhe ainda neste capítulo, pretendendo-se apenas com esta brevíssima introdução realçar a modernidade do trabalho de investigação nesta área da estatística em que o primeiro livro de texto cobrindo os casos que apresentámos, da autoria de Cressie (1991; 2ª edição em 1993), tem uns escassos 17 anos.

1.2 Objectivos da análise de dados espaciais

A diversidade de exemplos apresentados na introdução histórica permite perceber que os fenómenos, em que a localização no espaço físico onde se desenrolam tem um papel importante no seu desenvolvimento, surgem nos mais diversos ramos de actividade e são objecto de estudo em muitos campos de investigação, como por exemplo a climatologia, a ecologia, a epidemiologia, a sociologia ou a pesquisa agrícola ou mineira.

Em qualquer desses campos, os investigadores são cada vez mais confrontados com a tarefa de analisar dados que, para além de serem altamente multivariados e temporalmente correlacionados, estão também referenciados geograficamente e são muitas vezes apresentados em mapas. Normalmente, para esse tipo de dados, pretende-se modelar tendências e estruturas de correlação, estimar parâmetros dos modelos pressupostos, realizar testes de selecção a modelos que compitam entre si e, cada vez mais frequentemente, efectuar predição em instantes e locais não observados.

Neste curso faz-se uma primeira abordagem da análise destes conjuntos complexos de dados espaciais e, para melhor nos movermos neste vasto grupo, vamos classificá-los em três categorias sem aprofundarmos, para já, a definição dos modelos que se supõem estar subjacentes, mas realçando através de exemplos as suas diferenças e a investigação que mais frequentemente lhes está associada.

- **Dados referentes a pontos**

Sob esta designação são estudados conjuntos finitos de vectores, $\{\boldsymbol{y}(\boldsymbol{s}_1), \boldsymbol{y}(\boldsymbol{s}_2), \ldots, \boldsymbol{y}(\boldsymbol{s}_n)\}$, constituídos por observações das mesmas

características de um fenómeno em n localizações precisas de um domínio espacial contínuo limitado.

Neste curso, abordaremos apenas casos em que $\boldsymbol{y}(\boldsymbol{s})$ é unidimensional e em que o domínio está em $\mathbb{R}^2$, mas o estudo poderia ser facilmente estendido a espaços de dimensão superior.

Supostamente, o fenómeno poderia ser observado em todos os pontos desse domínio fixo, que notaremos por $\mathcal{D}$, e por essa razão pode ser modelado por um campo aleatório $\{Y(\boldsymbol{s}), \boldsymbol{s} \in \mathcal{D}\}$, em que $\mathcal{D}$ é um subconjunto de $\mathbb{R}^2$ contendo um rectângulo de área não nula.

Um exemplo de dados deste tipo é apresentado na Figura 1.3.

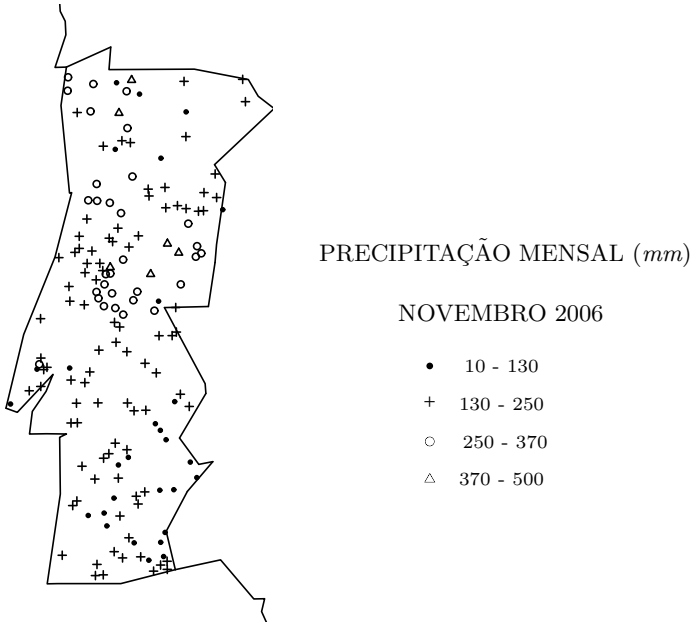


Figura 1.3 *Precipitação do mês de Novembro de 2006 registada em 158 estações da rede meteorológica nacional*

O fenómeno em análise é precipitação mensal em Portugal continental e os dados referem-se à precipitação acumulada, medida em milí-

metros, registada em 158 estações da rede meteorológica nacional durante o mês de Novembro de 2006 [1].

Embora a precipitação deste mês tenha sido forte, a sua distribuição está longe de ser uniforme, verificando-se acumulações particularmente intensas nas terras altas do centro e na faixa litoral norte do país, ao mesmo tempo que as terras do interior sul registavam valores bastante mais baixos.

O fim em vista neste exemplo é o de encontrar um modelo para a distribuição geográfica da precipitação que, tendo em conta a correlação espacial e algumas covariáveis como, por exemplo, a altitude, a distância ao mar, etc., possa fornecer estimativas em locais onde não existem estações meteorológicas.

Em dados associados a pontos, o propósito do estudo é, tal como neste exemplo, a predição do valor da característica analisada em pontos não observados. A modelação do campo aleatório subjacente ao fenómeno, a estimação dos respectivos parâmetros, a avaliação do ajustamento, são elementos essenciais na prossecução desse objectivo.

Habitualmente essa estimação é feita para uma grelha de pontos de malha apertada cobrindo o domínio, o que permite delimitar zonas com valores da variável em estudo dentro de intervalos pré-definidos, normalmente representadas em gráficos designados por **gráficos de contorno**, ou construir mapas que, associando tons de uma escala de cinzentos ou cores diferentes a valores da variável em intervalos que particionam o seu suporte, fornecem uma imagem da distribuição espacial do processo.

Como veremos no capítulo dedicado a análise de dados referentes a pontos, um dos elementos chave da modelação é o tratamento da

[1] Dados recolhidos na página electrónica do Instituto da Água (INAG), em Maio de 2007, < http://snirh.inag.pt> .

covariância entre variáveis associadas a duas localizações do domínio, a **covariância espacial**, que se supõe depender do vector que as une ou, em muitos casos, apenas do seu comprimento, a distância entre as localizações.

Saber calcular correctamente distâncias entre os pontos do domínio torna-se um imperativo e, uma vez que muitas aplicações dizem respeito a domínios na crusta terrestre, referiremos posteriormente de forma breve os problemas associados a este cálculo.

- **Dados referentes a áreas**

Esta designação utiliza-se para referir conjuntos de vectores, $\{\boldsymbol{y}(B_1), \boldsymbol{y}(B_2), \ldots, \boldsymbol{y}(B_n)\}$, constituídos por observações das mesmas características em n sub-regiões de uma partição de um domínio espacial fixo limitado.

Só serão consideradas observações unidimensionais em que o domínio espacial, $\mathcal{D}$, está contido em $\mathbb{R}^2$ e se encontra particionado num número finito de regiões regulares ou irregulares de área não nula com fronteiras bem definidas.

O campo aleatório Y usado para modelar este tipo de observações tem como espaço de índices a referida partição de $\mathcal{D}$, cujos elementos são designados por **células**, **áreas** ou **blocos**, sendo por isso notados por B_i em vez de $\boldsymbol{s}_i$.

Frequentemente, escolhe-se um ponto especial em cada bloco, muitas vezes o seu centróide, para o representar, e à partição do domínio associa-se um grafo, cujos vértices são exactamente os pontos representantes. Estes vértices estarão ligados por uma aresta se correspondem a células vizinhas, sendo que a relação de vizinhança entre células pode ser definida de muitas formas diferentes. A forma mais comum corresponde a considerar como vizinhas as células contíguas,

ou seja, as que partilham parte das respectivas fronteiras, mas a representação em grafo possibilita definições de vizinhança baseadas na distância entre vértices, que se tornam úteis em alguns tipos de estudos da correlação espacial entre células.

Na Figura 1.4 apresenta-se um exemplo da representação de dados deste tipo. Trata-se do número de casos de morte por cancro de estômago durante o período de 1991 a 1995, por concelho de residência [2]. Os dados aqui representados referem-se a toda a população, sem distinção de género ou classe etária, e abrangem apenas Portugal continental.

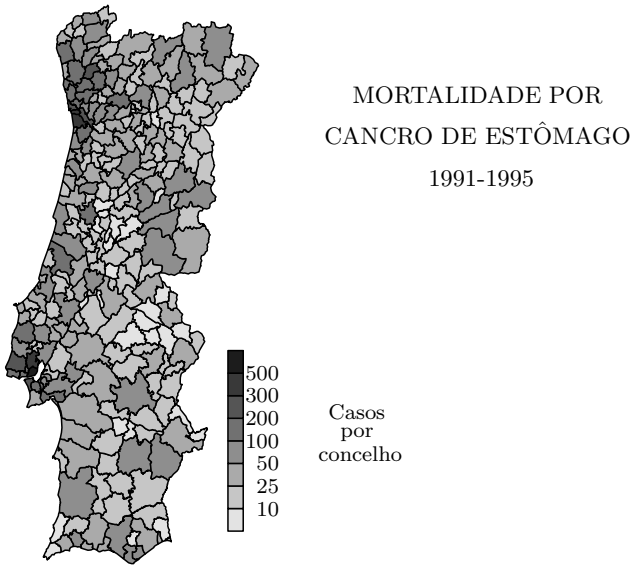


Figura 1.4 *Número de casos de morte de cancro de estômago por concelho de residência em Portugal continental no período de 1991 a 1995*

[2] Direcção Geral de Saúde.

O conhecimento dos locais exactos de residência dos casos de cancro aproximaria estes dados do tipo de dados envolvidos no estudo de Snow sobre a epidemia de cólera apresentado inicialmente, no entanto, as instituições produtoras de estatísticas oficiais só divulgam informação do número de casos até um determinado nível regional, o que transforma o estudo numa análise de dados areolares. De facto, os dados referentes a regiões administrativas dizem frequentemente respeito a fenómenos que envolvem alguma confidencialidade, como é o caso dos dados de saúde, e, como tal, só podem ser divulgados de forma agrupada sem risco de perda de anonimato.

Os mapas como o da figura anterior são chamados **mapas coropletos** ou **mapas de manchas** visto usarem manchas de cor para classificar os valores da variável em estudo num pequeno número de classes. Os mapas coropletos são cada vez mais utilizados por fornecerem uma ideia sumária, facilmente interpretável, da distribuição regional da característica de interesse. O reverso desta simplicidade é a falta de robustez da mensagem que transmitem, que se apoia essencialmente na sua natureza visual, facilmente manipulável por alteração dos limites das classes ou do gradiente de cores utilizado.

No presente exemplo estamos interessados em saber se existirá alguma tendência para os casos se aglomerarem em certas zonas do país, depois de controlados os efeitos óbvios que a concentração populacional nas zonas litorais e a acentuada diferença entre a distribuição etária dos habitantes dessas zonas e a dos do interior produzem no mapa apresentado.

A adopção de modelos que permitem construir estimativas alisadas do número de casos em cada região possibilita a obtenção de imagens mais realistas da incidência da doença e, inclusivamente, o despiste de causas que podem estar a determinar uma tão grande prevalência deste tipo de cancro em Portugal.

Tal como neste exemplo, a análise de dados referentes a áreas tem muitas vezes como objectivo a procura de padrões espaciais da distribuição da variável, nomeadamente a verificação da significância estatística de aglomerações de valores mais altos ou mais baixos em determinadas zonas do domínio, mas a predição de valores da variável em células não observadas é também um alvo habitual de estudos entre os quais se salientam os relacionados com o restauro de imagem.

Métodos baseados em modelos que levam em conta a correlação espacial entre células vizinhas e eventuais covariáveis são um instrumento precioso neste tipo de análise.

- **Dados referentes a processos pontuais**

Esta terceira classe de dados pode ser exemplificada pelas localizações no mapa da Figura 1.2 das moradas de casos de morte por cólera, ou pelas localizações exactas de árvores com certas características numa floresta.

Nos casos que trataremos, podem encarar-se os dados como conjuntos de localizações aleatórias correspondentes às ocorrências de um certo acontecimento de interesse num domínio fixo, W, de $\mathbb{R}^2$, designado por **janela**. Cada conjunto de localizações, **configuração**, é notado por $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ e serve de base ao processo, designado por **processo pontual**, usado na modelação.

Por vezes, a complementar cada ponto x das configurações do processo pontual, existe informação adicional sobre o tipo de ocorrência aí verificada, sob a forma de outra variável aleatória com valores num conjunto de marcas. Neste caso o processo chama-se **processo pontual marcado**.

A análise de dados referentes a processos pontuais tem frequentemente como objectivo o estudo do agrupamento das ocorrências no

espaço. Trata-se de verificar se as localizações tendem a estar aglomeradas espacialmente em um ou vários focos ou se, pelo contrário, se encontram espalhadas independentente e homogeneamente no espaço.

1.3 Modelos e realidade

Como em todos os casos em que se utilizam modelos matemáticos, deve ter-se presente que a adopção de um modelo para representar um certo conjunto de dados espaciais vai necessariamente condicionar a resposta aos diferentes problemas que se pretendem estudar e sobre os quais é frequentemente necessário tomar decisões, algumas delas com fortes implicações financeiras ou sociais.

Convém, por isso, realçar que os campos aleatórios são meramente modelos que se mostram úteis no tratamento de dados espaciais se a sua selecção for adequada e, nesse sentido, a primeira etapa da modelação é a escolha dum processo estocástico que tenha propriedades em consonância com as propriedades do fenómeno físico em estudo. Embora nunca se possa ter a certeza que o fenómeno tem as propriedades que atribuímos ao modelo, os dados não devem desmenti-las e, por isso, o ajustamento do modelo à realidade deve ser validado em todas as etapas de tratamento dos dados, para que se possa distinguir entre os resultados que são inerentes ao fenómeno e os que são produto do modelo que adoptámos.

Desde que feita com as precauções que apontámos, a adopção de modelos tem a vantagem de sumarizar as propriedades do fenómeno, o que facilita a distinção entre as que lhe são intrínsecas e as que são meramente fortuitas, sendo que nesse processo também têm o seu papel os modelos que foram afastados por se mostrarem inadequados.

Faz-se notar também que vários modelos podem estar concordantes

com as propriedades gerais manifestadas pelos dados e que, no caso de ser necessário seleccionar um deles, esta escolha deve ser guiada pelo problema científico subjacente ou, se este não fornecer informação adicional suficiente, pelo simples equilíbrio entre o ajustamento aos dados e a parcimónia do modelo.

Para finalizar, chamamos ainda a atenção que os modelos espaciais vêm responder à necessidade de alargar o conjunto dos métodos clássicos de análise de dados, provocada pela inadaptação da hipótese de independência das observações, ao tratamento de dados espaciais. De facto, é simplista supor que a única fonte de variabilidade deste tipo de dados são erros de medida modelados por um ruído branco, quando se torna tão evidente, tal como acontece com os fenómenos variando no tempo, que dados que estão mais perto no espaço têm tendência a ser mais parecidos dos que os que estão mais afastados. É na capacidade de incorporar, por diversos métodos, esta variação espacial, que estes modelos se distinguem dos clássicos.

1.3.1 Cálculo de distâncias sobre o globo terrestre

Em todos os exemplos que apresentámos esteve sempre presente a ideia que a dependência entre observações estava ligada à sua proximidade física, ou seja, que poderia ser traduzida numa função da distância entre localizações. Torna-se, por isso, imperativo saber determinar distâncias entre pontos do domínio.

Esse cálculo não envolve qualquer dificuldade nos fenómenos que são registados em superfícies planas, como por exemplo as ligadas à imagiologia médica ou de satélite, ou que dizem respeito a domínios terrestres pouco extensos. Para o fazer pode, por vezes, ser utilizado um qualquer instrumento de medida ou, no caso mais vulgar em que os pontos são dados em coordenadas cartesianas, através da distância euclidiana vulgar.

O mesmo não se pode dizer quando se trabalha com domínios terrestres maiores, como por exemplo o continente europeu, em que se torna necessário contar com a curvatura terrestre no cálculo de distâncias entre localizações geográficas mais distantes. Esta distância é o comprimento do caminho mais curto entre dois pontos sobre a superfície da Terra, corresponde ao comprimento do arco do círculo máximo entre os dois locais, e designa-se por **distância geodésica**.

Passemos a verificar como se calcula a distância geodésica quando, para qualquer localização $\boldsymbol{s}$, se conhecem as suas coordenadas geográficas, notadas por (λ, φ), em que λ representa a longitude (E-W), positiva para leste do meridiano de Greenwich, e φ a latitude (N-S), positiva para norte do equador.

No que se segue, vamos supor que estas coordenadas, dadas habitualmente em graus, minutos e segundos, já se encontram expressas em radianos através da expressão $\theta\,\pi/180$, em que θ é um ângulo expresso em graus decimalizados.

Consideremos, então, cada ponto representado em $\mathbb{R}^3$ pelas coordenadas polares num sistema de eixos ortogonais, (x, y, z), com origem no centro da Terra, eixo dos zz sobre o eixo de rotação da Terra e eixo dos xx no plano do equador unindo o centro da Terra ao meridiano de Greenwich. É simples verificar através da Figura 1.5 que, para um ponto $\boldsymbol{s}$ de coordenadas geográficas (λ, φ), as coordenadas polares são dadas por

$$\begin{cases} x = R\cos\varphi\cos\lambda \\ y = R\cos\varphi\,\mathrm{sen}\,\lambda \\ z = R\,\mathrm{sen}\,\varphi \end{cases}$$

em que R é o raio da terra, que se considera medir 6371 Km.

Consideremos agora dois pontos $\boldsymbol{s}_1 = (\lambda_1, \varphi_1)$ e $\boldsymbol{s}_2 = (\lambda_2, \varphi_2)$ sobre a

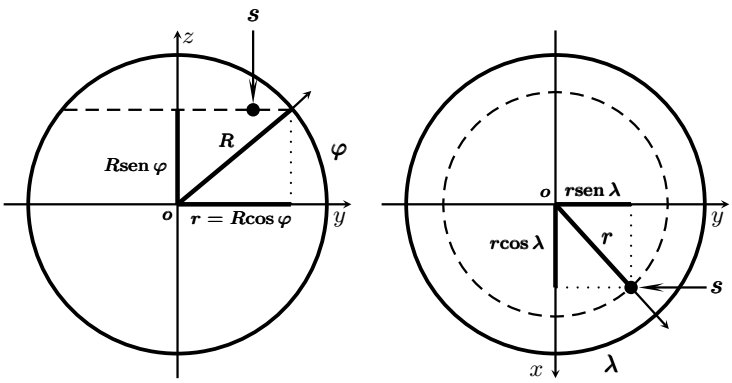


Figura 1.5 *Projecções dum ponto* $\boldsymbol{s} = (\lambda, \varphi)$ *do primeiro octante: à esquerda no plano yoz; à direita no plano xoy.*

superfície da terra, e seja θ o ângulo ao centro expresso em radianos correspondente ao arco do círculo máximo que os une. Temos então para a distância geodésica entre os dois pontos, $d(\boldsymbol{s}_1, \boldsymbol{s}_2) = R\theta$.

A melhor forma de calcular θ é através do seu coseno que é fácil de escrever em termos dos vectores $\boldsymbol{u}_1 = \overrightarrow{\boldsymbol{os}_1}$ e $\boldsymbol{u}_2 = \overrightarrow{\boldsymbol{os}_2}$ em que o ponto $\boldsymbol{o}$ é o centro da terra

$$\cos\theta = \frac{< \boldsymbol{u}_1.\boldsymbol{u}_2 >}{\|\boldsymbol{u}_1\|\|\boldsymbol{u}_2\|}.$$

Como

$$\begin{aligned}
< \boldsymbol{u}_1.\boldsymbol{u}_2 > &= x_1x_2 + y_1y_2 + z_1z_2 \\
&= R\cos\varphi_1\cos\lambda_1 R\cos\varphi_2\cos\lambda_2 \\
&+ R\cos\varphi_1 \operatorname{sen}\lambda_1 R\cos\varphi_2 \operatorname{sen}\lambda_2 + R\operatorname{sen}\varphi_1 R\operatorname{sen}\varphi_2 \\
&= R^2[\cos\varphi_1\cos\varphi_2(\cos\lambda_1\cos\lambda_2 + \operatorname{sen}\lambda_1\operatorname{sen}\lambda_2) \\
&+ \operatorname{sen}\varphi_1\operatorname{sen}\varphi_2] \\
&= R^2[\cos\varphi_1\cos\varphi_2\cos(\lambda_1 - \lambda_2) + \operatorname{sen}\varphi_1\operatorname{sen}\varphi_2],
\end{aligned}$$

e $\|\boldsymbol{u}_1\| = \|\boldsymbol{u}_2\| = R$, obtém-se

$$\cos\theta = \cos\varphi_1 \cos\varphi_2 \cos(\lambda_1 - \lambda_2) + \operatorname{sen}\varphi_1 \operatorname{sen}\varphi_2$$

e

$$d(\boldsymbol{s}_1, \boldsymbol{s}_2) = R \arccos[\cos\varphi_1 \cos\varphi_2 \cos(\lambda_1 - \lambda_2) + \operatorname{sen}\varphi_1 \operatorname{sen}\varphi_2].$$

Note-se que não é raro ver este tipo de cálculos trignométricos da distância geodésica substituídos por um cálculo em que se trata a longitude e a latitude como equivalentes a coordenadas cartesianas. Embora para pontos na vizinhança do equador o erro cometido não seja substancial, este aumenta muito quando se medem distâncias entre locais de longitude diferente que estão afastados do equador. De facto, o comprimento do arco correspondente a um grau de longitude diminui rapidamente com a aproximação aos pólos.

Esta questão prende-se com o problema mais geral da medida das distâncias em mapas que, por sua vez, está relacionado com a impossibilidade de preparar um mapa plano livre de distorção de uma superfície curva em todas as direcções, como acontece com a superfície da Terra. Como tal, quando se pretende construir uma representação sistemática de toda ou parte da superfície da Terra num plano, ou seja construir um mapa, tem que se escolher a característica, ou características, que se pretende que sejam mostradas com precisão e utilizar um tipo de projecção - função definida na superfície da Terra com valores no plano - que as preserve.

No caso vertente estaríamos interessados em projecções que preservassem as distâncias entre pontos. Infelizmente não existe, como Gauss demonstrou, uma projecção em mapa planar que preserve a distância entre todos os pontos e, como tal, procura remediar-se essa situação criando mapas que se dirigem a pequenas porções da globo terrestre para as quais a distorção não seja muito relevante.

Uma projecção que é habitualmente usada para esse efeito é a **projecção de Mercator transversa** em que se usa um cilindro com eixo perpendicular ao eixo de rotação da Terra e tangente à superfície desta num meridiano, para nele se fazer a projecção de cada ponto da Terra a partir do eixo do cilindro. Esta projecção, dita cilíndrica, é conforme (preserva os ângulos e as suas orientações), e produz uma distorção baixa numa faixa de 6° de longitude centrada no meridiano tangente. Normalmente essa faixa é ainda subdividida em 16 zonas com 8° de latitude que não incluem as regiões polares. Para pontos dentro de cada uma dessas zonas é possivel calcular as distâncias no mapa correspondente usando o teorema de Pitágoras visto que as unidades usadas são métricas.

Evidentemente que o primeiro passo na construção de um mapa deste tipo é a escolha de um modelo para a superfície da Terra. Em muitos casos é escolhida a forma esférica, embora a forma da Terra seja aproximadamente a de um elipsóide com o achatamento nos pólos. Seleccionar um destes dois modelos para representar a Terra implica escolher entre as vantagens e desvantagens de cada um deles. Para representações em pequena escala, como é o caso dos atlas e globos, escolhe-se normalmente o modelo esférico mas, para representações em escala grande ou média, escolhe-se o elipsóide. A forma esférica da projecção de Mercator transversa foi pela primeira vez apresentada por Lambert em 1772. Uma forma elíptica foi apresentada por Gauss em 1822 e analisada por Krüger no início do século XX sendo, por isso, chamada de **projecção de Gauss-Krüger** ou **projecção conforme de Gauss**.

Existe ainda um terceiro modelo mais complexo, chamado **geóide**, que representa o nível global médio do mar e que é obtido através de uma combinação de medidas terrestres e de satélite. Este modelo não é usado para fazer projecções em mapas mas sim para controlar

os **Datums Geográficos**.

Um datum geográfico é um sistema de referência constituído por: um ponto na superfície terrestre em relação ao qual as medidas de posição são feitas, um modelo da superfície da terra para calcular as posições, e um sistema de projecção. Historicamente os datums são baseados no elipsóide que melhor representa o geóide dentro da região que o datum pretende representar acrescentando-se-lhe algumas modificações (controlos) fornecidas pelo geóide.

Existe um elipsóide global de referência para representar toda a Terra cuja última versão é conhecida por WGS84 (World Geodetic System 1984), mas para a Europa é habitualmente usado o elipsóide de Hayford e para a América do Norte o elipsóide de Clarke.

Em Portugal existem vários datums [3], como por exemplo o Datum Lisboa, o Datum 73, etc., para o continente, o Datum Base SE-Porto Santo para a Madeira, e um Datum para cada um dos grupos do arquipélago dos Açores. Esses datums usam o elipsóide de Hayford para representar a terra e o sistema de projecção de Gauss-Kruger.

[3] Informação sobre os Datums usados em Portugal pode ser recolhida na página electrónica <http://www.igeo.pt> do Instituto Geográfico Português.

Note também que o pacote `mapproj` do software `R` possui a função `mapproject` que permite transformar coordenadas geográficas em coordenadas referentes a numerosos sistemas de projecção.

Capítulo 2

Modelos referentes a pontos

Neste capítulo vamos apresentar a análise clássica dos modelos para dados espaciais referentes a pontos também conhecida por **Geoestatística**.

2.1 Nomenclatura e conceitos básicos

Como foi dito anteriormente, os dados que analisaremos consistem em medidas de uma variável de interesse num conjunto finito de locais, $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\}$, pertencentes a um certo domínio de área não nula, $\mathcal{D} \subseteq \mathbb{R}^2$. Como foi também referido, este estudo poderia ser alargado a observações multidimensionais e a domínios incluídos em espaços de maior dimensão, fazendo alterações convenientes nos modelos que vamos utilizar.

2.1.1 Hipóteses iniciais

Cada observação $y_i = y(\boldsymbol{s}_i)$ será considerada como uma realização de uma variável aleatória $Y_i = Y(\boldsymbol{s}_i)$, cuja distribuição é governada por um processo estocástico subjacente, espacialmente contínuo, $\{Y(\boldsymbol{s}), \boldsymbol{s} \in \mathcal{D}\}$, notado simplesmente por Y quando não existir possibilidade de confusão, que é a restrição a $\mathcal{D}$ do processo em todo o plano $\{Y(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^2\}$, habitualmente designado por **campo**

aleatório.

Faz-se notar que os locais de observação do processo terão que ser escolhidos de acordo com planos de amostragem ditos **planos não preferenciais**, os únicos que admitiremos. Estes incluem dois tipos: determinístico (por exemplo, uma quadrícula uniforme de pontos cobrindo o domínio); estocástico, mas independente do processo Y.

Planos de amostragem com localizações dependentes da variável de observação são frequentes na vida real. Veja-se, por exemplo, a distribuição das estações de monitorização de poluentes atmosféricos, normalmente localizadas em pontos onde se espera que a poluição seja elevada, por forma a verificar-se o cumprimento das regras de concentração máxima de poluentes impostas pelas leis ambientais.

No que se segue, vamos admitir que diversas hipóteses são verificadas pelos processos espaciais subjacentes.

Notemos, respectivamente, por $\mathrm{E}[Y(\boldsymbol{s})] = \mu(\boldsymbol{s})$ e $\mathrm{Var}[Y(\boldsymbol{s})] = \sigma(\boldsymbol{s})$ a média e a variância do processo.

Hipótese 2.1. *O processo tem média e variância finitas,*

$$\mathrm{E}\big[|Y(\boldsymbol{s})|\big] < +\infty, \;\; \boldsymbol{s} \in \mathcal{D}, \qquad \mathrm{Var}\big(Y(\boldsymbol{s})\big) < +\infty, \;\; \boldsymbol{s} \in \mathcal{D}.$$

Hipóteses sobre a estacionaridade do processo são também admitidas e, por isso, vamos fazer uma breve revisão de alguns dos tipos que são habitualmente pressupostos.

2.1.2 Estacionaridade dos processos espaciais

Por vezes, consideram-se **processos espaciais fortemente estacionários**, ou seja, exige-se que, para todo o $n \geq 1$, para qualquer conjunto de locais $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\} \subset \mathcal{D}$ e qualquer vector $\boldsymbol{h}$ de $\mathbb{R}^2$ tal

que $\{\boldsymbol{s}_1+\boldsymbol{h}, \boldsymbol{s}_2+\boldsymbol{h}, \ldots, \boldsymbol{s}_n+\boldsymbol{h}\} \subset \mathcal{D}$, as distribuições de

$$\big(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n)\big) \quad \text{e} \quad \big(Y(\boldsymbol{s}_1+\boldsymbol{h}), Y(\boldsymbol{s}_2+\boldsymbol{h}), \ldots, Y(\boldsymbol{s}_n+\boldsymbol{h})\big)$$

sejam idênticas.

Usualmente consideram-se processos com estacionaridade menos restritiva, como por exemplo a estacionaridade de segunda ordem.

Um processo diz-se **estacionário de segunda ordem** ou fracamente estacionário se o processo é estacionário de primeira ordem,

$$E[Y(\boldsymbol{s})] = \mu, \;\; \boldsymbol{s} \in \mathcal{D}$$

e

$$\mathrm{Cov}\big(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)\big) = C(\boldsymbol{s}_1 - \boldsymbol{s}_2), \;\; \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{D},$$

ou seja, se a covariância depende apenas do vector de separação $\boldsymbol{s}_1 - \boldsymbol{s}_2$.

A função $C(.)$ é conhecida como **função de covariância** ou **covariograma**.

Note-se que esta definição implica que o processo tenha variância constante, $\sigma^2(\boldsymbol{s}) = C(\boldsymbol{0})$, e ainda que o covariograma seja simétrico, $C(\boldsymbol{h}) = C(-\boldsymbol{h})$.
Se $C(\boldsymbol{0}) \neq 0$ é equivalente definir a segunda propriedade à custa da **função de correlação** ou **correlograma**[1], notada por $\rho(.)$,

$$\mathrm{Corr}\big(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)\big) = \frac{\mathrm{Cov}\big(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)\big)}{C(\boldsymbol{0})} = \rho(\boldsymbol{s}_1 - \boldsymbol{s}_2), \quad (2.1)$$

para a qual se verifica $\rho(\boldsymbol{h}) = \rho(-\boldsymbol{h})$ e $\rho(\boldsymbol{0}) = 1$.

Claro que, devido à hipótese de existência de variância que estabelecemos no início, se pode garantir que um processo fortemente estacionário também é fracamente estacionário, mas a afirmação recíproca não é verdadeira.

[1] Alguns autores usam as designações de covariograma e correlograma exclusivamente para os gráficos da função de covariância e de correlação.

Por vezes, exige-se um tipo de estacionaridade ainda mais fraco.

Um processo diz-se **intrinsecamente estacionário** se

$$\mathrm{E}\left[Y(\boldsymbol{s}_1) - Y(\boldsymbol{s}_2)\right] = 0, \quad \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{D} \tag{2.2}$$

e

$$\begin{aligned} 2\gamma\big(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)\big) &= \mathrm{Var}\big(Y(\boldsymbol{s}_1) - Y(\boldsymbol{s}_2)\big) && (2.3)\\ &= \mathrm{E}\left[\big(Y(\boldsymbol{s}_1) - Y(\boldsymbol{s}_2)\big)^2\right] \\ &= 2\gamma\big(\boldsymbol{s}_1 - \boldsymbol{s}_2\big), \quad \boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathcal{D}. && (2.4) \end{aligned}$$

À função $2\gamma(.)$ chama-se **variograma** e a $\gamma(.)$ **semivariograma**.

Sem a Hipótese 2.1, o facto de um processo ser intrinsecamente estacionário não garantiria a existência de valor médio ou variância, no entanto, como se adoptou esta hipótese, a estacionaridade intrínseca obriga a que o processo tenha valor médio constante, i.e., que seja estacionário de primeira ordem. De facto, a estacionaridade intrínseca é uma propriedade associada a diferenças e, como tal, nada diz sobre a distribuição conjunta das observações $Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)$, ou seja, não proporciona o conhecimento da verosimilhança [2].

Evidentemente que existem relações entre a estacionaridade de segunda ordem e a estacionaridade intrínseca que convém conhecer.

É fácil verificar que

$$\begin{aligned} 2\gamma\big(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)\big) &= \mathrm{Var}\big(Y(\boldsymbol{s}_1) - Y(\boldsymbol{s}_2)\big) \\ &= \mathrm{Var}\big(Y(\boldsymbol{s}_1)\big) + \mathrm{Var}\big(Y(\boldsymbol{s}_2)\big) - 2\mathrm{Cov}\big(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)\big), \end{aligned}$$

[2] Note-se que, se a caracterização de um processo é feita à custa de diferenças, não se consegue distinguir entre os processos $Y(\boldsymbol{s}) = \mu + Z(\boldsymbol{s})$ e $Y'(\boldsymbol{s}) = M + Z(\boldsymbol{s})$, em que M é uma variável aleatória qualquer.

Se, no segundo caso, a variável M não tiver valor médio e Z for um processo com valor médio constante, Y é um exemplo de um processo sem valor médio e com incrementos satisfazendo a condição (2.2).

e esta expressão mostra que, conhecida a variância do processo, a adopção de um modelo de covariância determina um modelo para o variograma e reciprocamente.

O conhecimento da variância do processo é essencial nesta equivalência. Note-se que, se um processo é estacionário de segunda ordem, a variância é conhecida e constante, $\sigma^2(\boldsymbol{s}) = C(\mathbf{0})$, e vem

$$\begin{aligned} 2\gamma\big(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)\big) &= 2\left(C(\mathbf{0}) - \mathrm{Cov}\big(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)\big)\right) \\ \gamma\big(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)\big) &= C(\mathbf{0}) - \mathrm{Cov}\big(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)\big), \\ \gamma\big(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)\big) &= C(\mathbf{0}) - C(\boldsymbol{s}_1 - \boldsymbol{s}_2) = \gamma(\boldsymbol{s}_1 - \boldsymbol{s}_2) \end{aligned} \tag{2.5}$$

ou seja, um processo estacionário de segunda ordem é intrinsecamente estacionário e

$$\gamma(\boldsymbol{h}) = C(\mathbf{0}) - C(\boldsymbol{h}). \tag{2.6}$$

O mesmo já não se pode dizer do recíproco. Mesmo sabendo que o processo é estacionário de segunda ordem, se se conhecer exclusivamente $\gamma(\boldsymbol{h})$, a covariância fica conhecida a menos de uma constante. No entanto, se também for válida a seguinte condição [3],

$$\mathrm{Cov}\big(Y(\boldsymbol{s}), Y(\boldsymbol{s}+\boldsymbol{h})\big) \to 0 \quad \text{quando } \|\boldsymbol{h}\| \to \infty \tag{2.7}$$

teremos, se aplicarmos limites a ambos os membros de (2.5),

$$\begin{aligned} \lim_{\|\boldsymbol{h}\|\to\infty} \gamma\left(Y(\boldsymbol{s}), Y(\boldsymbol{s}+\boldsymbol{h})\right) &= C(\mathbf{0}) \\ &\quad - \lim_{\|\boldsymbol{h}\|\to\infty} \mathrm{Cov}\big(Y(\boldsymbol{s}), Y(\boldsymbol{s}+\boldsymbol{h})\big) \\ \lim_{\|\boldsymbol{h}\|\to\infty} \gamma\big(Y(\boldsymbol{s}), Y(\boldsymbol{s}+\boldsymbol{h})\big) &= C(\mathbf{0}), \qquad \boldsymbol{s} \in \mathcal{D}, \end{aligned} \tag{2.8}$$

[3] De facto, se se exigir uma forma fraca de independência assintótica que é equivalente à ergodicidade, a condição 2.7 é verificada. Num processo ergódico esperanças respeitantes ao espaço de probabilidades base do processo puderem ser estimadas por médias espaciais. A classe dos processos ergódicos está contida estritamente na classe dos processos fortemente estacionários.

e poderemos escrever a covariância exclusivamente à custa de $\gamma(.)$,

$$\begin{aligned}\mathrm{Cov}\big(Y(\boldsymbol{s}), Y(\boldsymbol{s}+\boldsymbol{h})\big) &= C(\mathbf{0}) - \gamma\big(Y(\boldsymbol{s}), Y(\boldsymbol{s}+\boldsymbol{h})\big) \\ &= \lim_{\|\mathbf{u}\|\to\infty} \gamma\big(Y(\boldsymbol{s}), Y(\boldsymbol{s}+\mathbf{u})\big) + \gamma(\boldsymbol{h}) \\ &= \lim_{\|\mathbf{u}\|\to\infty} \gamma(\mathbf{u}) + \gamma(\boldsymbol{h}), \qquad \boldsymbol{s} \in \mathcal{D}.\end{aligned} \tag{2.9}$$

A expressão anterior mostra que um processo intrinsecamente estacionário para o qual seja válido,

$$\lim_{\|\mathbf{u}\|\to\infty} \gamma(\mathbf{u}) < \infty, \tag{2.10}$$

é um processo estacionário de segunda ordem com covariograma dado por

$$C(\boldsymbol{h}) = \lim_{\|\mathbf{u}\|\to\infty} \gamma(\mathbf{u}) + \gamma(\boldsymbol{h}). \tag{2.11}$$

Estas conclusões mostram que a inclusão dos processos de segunda ordem na classe dos processos intrinsecamente estacionários é uma inclusão estrita. De facto, verificaremos mais à frente que existem variogramas válidos para os quais o limite (2.10) é infinito.

2.1.3 Propriedades do variograma e covariograma

Passamos a estudar mais em pormenor as propriedades do semivariograma de um processo estacionário de segunda ordem.

Deduz-se facilmente da definição que $\gamma(-\boldsymbol{h}) = \gamma(\boldsymbol{h}),\ \gamma(\mathbf{0}) = 0$.

Admite-se, no entanto, que possa acontecer,

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

em que τ^2 é chamado o **efeito de pepita**.

Esta descontinuidade do semivariograma na origem é teoricamente impossível se o processo a que corresponde for contínuo em média

quadrática, ou seja, se $E[(Y(\boldsymbol{s}+\boldsymbol{h})-Y(\boldsymbol{s}))^2] \to 0$, quando $\|\boldsymbol{h}\| \to 0$, mas ela foi precisamente admitida para modelar os fenómenos que têm um comportamento descontínuo em micro-escala, ou seja, aqueles para os quais se pode supor que em cada ponto existe uma pequeníssima pepita do mesmo material que é diferente do material vizinho.

A presença da descontinuidade pode também ser explicada através da decomposição do processo em vários subprocessos cada um deles reflectindo o comportamento do fenómeno numa determinada escala:
- Uma primeira componente determinística, $\mu(.)$, convencionalmente designada em geoestatística por **tendência espacial**, reflecte o comportamento médio do processo e traduz a variabilidade em larga escala;
- Uma segunda componente, $W(.)$, de valor médio nulo, estacionária de segunda ordem e contínua em média quadrática, modela a ligação espacial entre as observações e traduz a variabilidade em pequena e micro-escala [4];
- Finalmente, uma terceira componente pretende representar os erros de medida e é modelada por um ruído branco, $\epsilon(.)$, independente da componente anterior, que deriva do pressuposto que, se fossem feitas várias observações do processo numa certa localização, estas flutuariam em torno do correspondente valor do processo.

[4] Por vezes, esta parcela da decomposição é ainda decomposta em duas $W(\boldsymbol{s}) = W_1(\boldsymbol{s}) + W_2(\boldsymbol{s})$, estacionárias de segunda ordem e independentes. $W_1(.)$ explica a variação em pequena escala, (entre pontos $\boldsymbol{s}_i, \boldsymbol{s}_j$ tais que $\|\boldsymbol{s}_i - \boldsymbol{s}_j\| \geq \min \|\boldsymbol{s}_k - \boldsymbol{s}_l\|$, $(1 \leq k, l \leq n)$, e $W_2(.)$ explica a variação em micro-escala, (entre pontos $\boldsymbol{s}_i, \boldsymbol{s}_j$ tais que $\|\boldsymbol{s}_i - \boldsymbol{s}_j\| < \min \|\boldsymbol{s}_k - \boldsymbol{s}_l\|$, $(1 \leq k, l \leq n)$. Na prática, a não ser que existam observações repetidas nas mesma localizações, é impossível distinguir entre o efeito de $W_2(.)$ e de $\epsilon(.)$ e, por essa razão, agrupam-se frequentemente numa só componente $\xi(.) = W_2(.) + \epsilon(.)$.

Suponhamos que o processo admite uma tal decomposição,

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + W(\boldsymbol{s}) + \epsilon(\boldsymbol{s}), \tag{2.12}$$

e seja $\mathrm{Var}\big(\epsilon(\boldsymbol{s})\big) = \tau^2$.

É fácil verificar as seguintes relações,

$$\begin{aligned} 2\gamma(\boldsymbol{h}) &= \mathrm{Var}\big(\mu(\boldsymbol{s}) + W(\boldsymbol{s}) + \epsilon(\boldsymbol{s}) - \mu(\boldsymbol{s}+\boldsymbol{h}) - W(\boldsymbol{s}+\boldsymbol{h}) - \epsilon(\boldsymbol{s}+\boldsymbol{h})\big) \\ &= \mathrm{Var}\big(W(\boldsymbol{s}) - W(\boldsymbol{s}+\boldsymbol{h})\big) + \mathrm{Var}\big(\epsilon(\boldsymbol{s}) - \epsilon(\boldsymbol{s}+\boldsymbol{h})\big), \end{aligned}$$

e ainda, dada a independência de $W(.)$, contínuo em média quadrática, e $\epsilon(.)$, ruído branco,

$$\begin{aligned} 2 \lim_{\|\boldsymbol{h}\|\to\boldsymbol{0}} \gamma(\boldsymbol{h}) &= \lim_{\|\boldsymbol{h}\|\to\boldsymbol{0}} E\big[\big(W(\boldsymbol{s}) - W(\boldsymbol{s}+\boldsymbol{h})\big)^2\big] \\ &\quad + \lim_{\|\boldsymbol{h}\|\to\boldsymbol{0}} \big\{\mathrm{Var}\big(\epsilon(\boldsymbol{s})\big) + \mathrm{Var}(\epsilon\big(\boldsymbol{s}+\boldsymbol{h})\big)\big\} = 2\tau^2 \neq 0, \end{aligned}$$

que, como pretendíamos, mostra como este modelo explica a descontinuidade do variograma na origem.

Usando as designações,

$$\mathrm{Var}(W(\boldsymbol{s})) = \sigma^2 \ \text{ e } \ \rho\big(W(\boldsymbol{s}), W(\boldsymbol{s}')\big) = \rho_W(\boldsymbol{s}-\boldsymbol{s}'),$$

temos ainda,

$$\mathrm{Var}\big(Y(\boldsymbol{s})\big) = C(\boldsymbol{0}) = \mathrm{Var}\big(W(\boldsymbol{s}) + \epsilon(\boldsymbol{s})\big) = \sigma^2 + \tau^2 \tag{2.13}$$

e

$$C(\boldsymbol{s}-\boldsymbol{s}') = \sigma^2 \rho_W(\boldsymbol{s}-\boldsymbol{s}'), \quad \boldsymbol{s} \neq \boldsymbol{s}',$$

o que evidencia que o covariograma será igualmente descontínuo na origem.

Note-se também que,

$$\rho(\boldsymbol{s}-\boldsymbol{s}') = \frac{C(\boldsymbol{s}-\boldsymbol{s}')}{C(\boldsymbol{0})} = \frac{\sigma^2}{\tau^2+\sigma^2}\rho_W(\boldsymbol{s}-\boldsymbol{s}')$$

o que conduz, através de (2.6), às seguintes relações para o semivariograma,

$$\begin{aligned}\gamma(\boldsymbol{s}-\boldsymbol{s}') &= C(\mathbf{0}) - C(\mathbf{0})\rho(\boldsymbol{s}-\boldsymbol{s}') \\ &= (\tau^2+\sigma^2)\big(1-\rho(\boldsymbol{s}-\boldsymbol{s}')\big),\end{aligned} \tag{2.14}$$

$$\begin{aligned}\gamma(\boldsymbol{s}-\boldsymbol{s}') &= (\tau^2+\sigma^2) - (\tau^2+\sigma^2)\big(\frac{\sigma^2}{\tau^2+\sigma^2}\rho_W(\boldsymbol{s}-\boldsymbol{s}')\big) \\ &= \tau^2+\sigma^2-\sigma^2\rho_W(\boldsymbol{s}-\boldsymbol{s}') \\ &= \tau^2+\sigma^2\big(1-\rho_W(\boldsymbol{s}-\boldsymbol{s}')\big).\end{aligned}$$

Devido a este último resultado, os semivariogramas de processos espaciais estacionários de segunda ordem são genericamente representados pela expressão

$$\gamma(\boldsymbol{h}) = \tau^2+\sigma^2\big(1-\rho'(\boldsymbol{h})\big), \tag{2.15}$$

em que $\rho'(.)$ é a função de correlação de um processo estacionário de segunda ordem, contínuo em média quadrática, e τ^2 e σ^2 são parâmetros não negativos.

De facto, na bibliografia mais recente, a função de correlação dos processos estacionários de segunda ordem passou a ser identificada com a função de correlação da componente W da decomposição (2.12), ou seja,

$$\rho(\boldsymbol{s}-\boldsymbol{s}') = \frac{C(\boldsymbol{s}-\boldsymbol{s}')}{\sigma^2},$$

em vez de $\rho(\boldsymbol{s}-\boldsymbol{s}') = C(\boldsymbol{s}-\boldsymbol{s}')/C(\mathbf{0})$, como definido em (2.1).

Para terminar a nomenclatura associada ao variograma, referimos que se utiliza a designação de **patamar** para

$$\lim_{\|\boldsymbol{h}\|\to\infty}\gamma(\boldsymbol{h}) = C(\mathbf{0})$$

quando este limite existe[5], e o de **patamar parcial** para a diferença entre o patamar e o efeito de pepita

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

O patamar parcial é notado por σ^2 o que leva a que, tal como em (2.13),

$$\lim_{\|\boldsymbol{h}\| \to \infty} \gamma(\boldsymbol{h}) = \tau^2 + \sigma^2. \tag{2.16}$$

De facto o patamar parcial pode ser identificado com a variância da componente W da decomposição (2.12).

Alguns autores chamam **alcance** ou **amplitude do variograma na direcção $\boldsymbol{h}/\|\boldsymbol{h}\|$** ao menor valor de $\|h\|$ para o qual se verifica

$$\gamma\big((1+\delta)\boldsymbol{h}\big) = C(\mathbf{0}),\ \delta > 0,$$

ou seja, à menor distância na direcção $\boldsymbol{h}/\|\boldsymbol{h}\|$ para a qual o semivariograma passa a ser constantemente igual ao patamar.

Vamos agora verificar uma propriedade importante do variograma que permite limitar o conjunto das funções que podem ser variogramas válidos.

Proposição 2.1. *Para qualquer inteiro $n > 0$, qualquer conjunto de localizações $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ e qualquer conjunto de constantes $\{a_1, \ldots, a_n\}$, tais que $\sum_{i=0}^{n} a_i = 0$, tem-se, para um semivariograma válido, a seguinte desigualdade*

$$\sum_i \sum_j a_i a_j \gamma\big(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\big) \leq 0,$$

o que é equivalente a dizer que a matriz $\Gamma = \Big(\gamma\big(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\big)\Big)$ é semidefinida negativa.

[5]Recorde-se que este limite pode não existir se o processo for apenas intrinsecamente estacionário.

Demonstração:

Note-se que

$$\sum_i \sum_j a_i a_j \gamma\big(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\big) = \frac{1}{2} \sum_i \sum_j a_i a_j E\big[(Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j))^2\big]$$
$$= \frac{1}{2} E\Big[\sum_i \sum_j a_i a_j (Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j))^2\Big]$$

$$= \frac{1}{2} E\Big[\sum_i \sum_j a_i a_j Y(\boldsymbol{s}_i)^2 + \sum_i \sum_j a_i a_j Y(\boldsymbol{s}_j)^2 - 2 \sum_i \sum_j a_i a_j Y(\boldsymbol{s}_i) Y(\boldsymbol{s}_j)\Big]$$
$$= \frac{1}{2} E\Big[\sum_i a_j Y(\boldsymbol{s}_i)^2 \underbrace{\sum_j a_i}_{0} + \sum_j a_j Y(\boldsymbol{s}_j)^2 \underbrace{\sum_i a_i}_{0} - 2 \sum_i \sum_j a_i a_j Y(\boldsymbol{s}_i) Y(\boldsymbol{s}_j)\Big]$$
$$= -E\Big[\sum_i a_i Y(\boldsymbol{s}_i) \sum_j a_j Y(\boldsymbol{s}_j)\Big] = -E\Big[\Big(\sum_i a_i Y(\boldsymbol{s}_i)\Big)^2\Big] \leq 0.$$

■

O covariograma tem propriedades análogas às do variograma, que se deduzem facilmente das deste último atendendo à relação já estabelecida

$$C(\boldsymbol{h}) = C(\boldsymbol{0}) - \gamma(\boldsymbol{h}).$$

Assim, usando a notação adoptada para o variograma, temos para o covariograma,

$$C(\boldsymbol{0}) = \tau^2 + \sigma^2, \quad \lim_{\|\boldsymbol{h}\| \to \boldsymbol{0}} C(\boldsymbol{h}) = \sigma^2 \quad \text{e} \quad \lim_{\|\boldsymbol{h}\| \to \infty} C(\boldsymbol{h}) = 0.$$

A demonstração da propriedade seguinte que estabelece que um covariograma válido tem de ser uma função definida positiva, obtém-se de imediato da Proposição 2.1, usando a relação (2.9).

Proposição 2.2. *Para qualquer inteiro $n > 0$, qualquer conjunto de localizações $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ e qualquer conjunto de constantes*

$\{a_i : a_i \neq 0,\ i = 1, \ldots, n\}$, tais que $\sum_{i=0}^{n} a_i = 0$, tem-se para um covariograma válido a seguinte desigualdade

$$\sum_i \sum_j a_i a_j \mathrm{Cov}\big(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\big) = \mathrm{Var}\Big(\sum_i a_i Y(\boldsymbol{s}_i)\Big) > 0,$$

ou seja, a matriz de covariâncias $\Sigma = \Big(\mathrm{Cov}\big(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\big)\Big)$ é definida positiva.

A utilização da propriedade anterior para averiguar a validade dum covariograma não é fácil. Existem, no entanto, condições necessárias e suficientes mais simples de verificar como, por exemplo, a dada pelo teorema de Bochner que, para campos aleatórios reais, diz que $C(\boldsymbol{h})$ é definida positiva se e só se

$$C(\boldsymbol{h}) = \int cos(\boldsymbol{w}^T \boldsymbol{h}) G(d\boldsymbol{w}), \tag{2.17}$$

onde G é uma medida em $\mathbb{R}^2$, limitada, positiva e simétrica em torno de $\mathbf{0}$.

Por outras palavras, $C(\boldsymbol{h})$ será uma função de covariância válida se e só se tiver representação espectral (2.17), em que $G/C(\mathbf{0})$, com $C(\mathbf{0}) = \int G(d\boldsymbol{w})$, é referida como a **distribuição espectral** que induz C. Se G tem densidade relativamente à medida de Lebesgue, i.e. $G(d\boldsymbol{w}) = g(\boldsymbol{w})d\boldsymbol{w}$, então $g(\boldsymbol{w})/C(\mathbf{0})$ é designada por **densidade espectral**.

Suponhamos, sem perda de generalidade, que existia densidade espectral. Uma vez que G é simétrica, (2.17) é equivalente a

$$C(\boldsymbol{h}) = \int e^{i\boldsymbol{w}^T \boldsymbol{h}} g(\boldsymbol{w}) d\boldsymbol{w}, \tag{2.18}$$

o que mostra que teremos uma função de covariância válida se e só se a função de correlação $\rho(\boldsymbol{h}) = C(\boldsymbol{h})/C(\mathbf{0})$ for a função característica de uma variável aleatória bidimensional com densidade simétrica $g/C(\mathbf{0})$.

Atente-se na transformada de Fourier de $C(\boldsymbol{h})$,

$$C^*(\boldsymbol{w}) = \frac{1}{2\pi} \int e^{-i\boldsymbol{w}^T\boldsymbol{h}} C(\boldsymbol{h})\ d(\boldsymbol{h}), \tag{2.19}$$

e também na fórmula de inversão desta transformada que, neste caso, resulta em

$$C(\boldsymbol{h}) = \frac{1}{2\pi} \int e^{i\boldsymbol{w}^T\boldsymbol{h}} C^*(\boldsymbol{w})\, d(\boldsymbol{w}).$$

Comparando esta última expressão com (2.18), conclui-se que $C^*(\boldsymbol{w})/2\pi = g(\boldsymbol{w})$ e, consequentemente, que

$$\int \frac{C^*(\boldsymbol{w})}{2\pi C(\boldsymbol{0})}\ d\boldsymbol{w} = 1.$$

Desta forma, pode verificar-se se uma certa função $C(\boldsymbol{h})$ é um covariograma válido calculando a sua transformada de Fourier (2.19) e verificando apenas se é positiva e integrável.

O cálculo da transformada de Fourier, excepto para os casos especiais em que se pode obter explicitamente, pode ser feito de forma aproximada através do eficiente algoritmo conhecido por FFT (Fast Fourier Transform) que, tendo sido desenvolvido para o cálculo da transformada de Fourier discreta, se adapta bem ao cálculo da transformada contínua.

Das propriedades das funções características retira-se imediatamente que o produto, bem como a combinação linear convexa de funções de covariância, é ainda uma função de covariância. Da mesma forma, a convolução de funções de covariância é uma função de covariância.

2.1.4 Isotropia e Anisotropia

A isotropia de um processo espacial define-se usualmente à custa do semivariograma. Assim, diz-se que um processo espacial intrinsecamente estacionário é **isotrópico** se o semivariograma depende de $\boldsymbol{h}$

apenas através do seu comprimento $\|\boldsymbol{h}\|$, ou seja, se quaisquer que sejam $\boldsymbol{s} \in \mathcal{D}$ e $\boldsymbol{h} \in \mathbb{R}^2$ tais que $s + \boldsymbol{h} \in \mathcal{D}$, se tem,

$$\gamma\big(Y(\boldsymbol{s}), Y(\boldsymbol{s}+h)\big) = \gamma(\|\boldsymbol{h}\|).$$

Um processo intrinsecamente estacionário isotrópico chama-se **homogéneo** e um processo que não é isotrópico diz-se **anisotrópico**.

Se o processo é também estacionário de segunda ordem, a isotropia pode ser definida por condição análoga para o covariograma,

$$\boldsymbol{s} \in \mathcal{D},\, \boldsymbol{h} \in \mathbb{R}^2 \text{ tais que } s + \boldsymbol{h} \in \mathcal{D}:\ \mathrm{Cov}\big(Y(\boldsymbol{s}), Y(\boldsymbol{s}+h)\big) = C(\|\boldsymbol{h}\|).$$

Note-se que, no caso de existir isotropia, os alcances direccionais do variograma atrás definidos têm igual valor. Desta forma, designa-se por **alcance** do variograma, e nota-se por r, a menor distância a partir da qual o variograma é igual ao patamar,

$$r = \min_t \gamma((1+\delta)t) = C(0),\ \delta > 0.$$

Uma vez que existem variogramas para os quais o patamar só é atingido assintoticamente define-se o **alcance efectivo** ou **alcance prático** como sendo a menor distância r_0 a partir da qual $\gamma(t) \geq 0.95(\tau^2+\sigma^2)$ para $t \geq r_0$. Evidentemente que esta definição pressupõe que nos estamos a restringir a variogramas estritamente monótonos.

Em processos estacionários de segunda ordem, usando a versão isotrópica da expressão (2.14) pode escrever-se a desigualdade anterior em função do correlograma,

$$\gamma(t) = (\tau^2 + \sigma^2)(1 - \rho(t)) \geq 0.95(\tau^2 + \sigma^2),$$

o que define o patamar efectivo como sendo a menor distância r_0 a partir da qual se tem

$$\rho(t) \leq 0.05,\ t \geq r_0,$$

ou seja, a menor distância a partir da qual a associação espacial se torna negligenciável.

Os autores que definem ρ', dado em (2.15), como correlograma do processo chamam alcance efectivo ao valor de r_0 a partir do qual se tem,

$$\rho'(t) \leq 0.05,\ t \geq r_0, \tag{2.20}$$

o que conduz a uma definição diferente da apresentada em termos do variograma,

$$\gamma(t) = \tau^2 + 0.95\sigma^2, \quad t \geq r_0.$$

Passaremos a adoptar esta última definição de alcance efectivo por ser a utilizada na bibliografia mais recente.

Passemos agora a tratar, embora brevemente, o fenómeno da anisotropia, que se percebe ser causado por comportamentos diferentes do processo segundo as direcções do espaço. Consoante o tipo de diferenças observadas, assim se podem definir diversos tipos de anisotropia. Comecemos pela mais frequentemente encontrada na vida real, conhecida por **anisotropia geométrica** que diz respeito à variação direccional do alcance do variograma.

Consideremos processos intrinsecamente estacionários. Um processo espacial diz-se com anisotropia geométrica se o seu variograma tem a seguinte forma:

$$\gamma(\boldsymbol{h}) = \gamma'\big(\|A\,\boldsymbol{h}\|\big),\ \boldsymbol{h} \in \mathbb{R}^2 \tag{2.21}$$

em que A é uma matriz invertível e $\gamma'(.)$ é um variograma isotrópico válido.

Os contornos correspondentes a valores constantes dum variograma deste tipo no sistema de eixos coordenados original são elipses. Todas estas elipses têm os eixos com a mesma orientação, bem como

o mesmo valor para a razão, $\lambda > 1$, entre o eixo principal (direcção segundo a qual o variograma cresce mais lentamente) e o eixo menor (direcção segundo a qual o variograma cresce mais rapidamente) designado por **factor de anisotropia**. O ângulo ψ do eixo principal com o eixo coordenado vertical, medido convencionalmente no sentido directo, designa-se por **ângulo de anisotropia**.

Note-se que em (2.21), a matriz A define uma aplicação linear do espaço original em si próprio que transforma as elipses em círculos. Através de uma rotação um novo conjunto de eixos é alinhado com o eixo principal e o eixo menor da elipse. Seguidamente um reescalonamento alonga o eixo menor de forma a que o seu comprimento iguale o do eixo principal.

Se $\theta = \frac{\pi}{2} - \psi$, então A é dado por

$$A = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{pmatrix}.$$

Assim, as coordenadas transformadas de um vector $\boldsymbol{h}$ de coordenadas originais (h_1, h_2) serão dadas por

$$\boldsymbol{h}' = \begin{pmatrix} h_1' \\ h_2' \end{pmatrix} = A \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

e $\gamma(\boldsymbol{h}) = \gamma'\big(\sqrt{h_1'^2 + h_2'^2}\,\big)$.

Funções de covariância com anisotropia geométrica podem definir-se analogamente,

$$C(\boldsymbol{h}) = \sigma^2 \rho\big((\boldsymbol{h}^T B \boldsymbol{h})^{1/2}\big),\ \boldsymbol{h} \neq \boldsymbol{0}$$

com B matriz definida positiva e $\rho(.)$ um correlograma isotrópico válido.

O conceito de anisotropia geométrica pode ser estendido para o de **anisotropia de alcance** utilizando uma transformação regular do

espaço que permita que os contornos do variograma possam ser mais gerais.

Os outros tipos de anisotropia surgem quando se deixa variar com a direcção do vector de separação, os outros dois parâmetros do variograma já definidos.

A anisotropia diz-se **anisotropia de pepita** se $\lim_{c\to 0}\gamma\Big(c\frac{\boldsymbol{h}}{\|\boldsymbol{h}\|}\Big)$ depende de $\boldsymbol{h}$, e diz-se **anisotropia de patamar** se $\lim_{c\to\infty}\gamma\Big(c\frac{\boldsymbol{h}}{\|\boldsymbol{h}\|}\Big)$ depende de $\boldsymbol{h}$. Esta última anisotropia pode configurar uma falha na estacionaridade de primeira ordem do processo.

2.1.5 Correlograma e diferenciabilidade do processo

A partir deste momento, passaremos a utilizar a versão isotrópica da expressão genérica (2.15),

$$\gamma(t) = \tau^2 + \sigma^2\big(1 - \rho(t)\big),$$

para referir o variograma de um processo estacionário de segunda ordem isotrópico. Pelas razões que já apontámos a função de correlação do processo passa também a ser definida por $\rho(t) = C(t)/\sigma^2$.

A utilidade desta representação fica bem patente no facto da lisura das trajectórias do processo espacial serem determinadas pelo correlograma. De facto, as trajectórias do processo podem ser encaradas como superfícies aleatórias sobre a região amostrada e, se se pretende fazer predição em pontos não amostrados, entre as características do processo que convém conhecer estarão necessariamente a continuidade e a diferenciabilidade destas superfícies.

Como se sabe a continuidade e a diferenciabilidade não têm, em processos aleatórios, uma única definição. Estas dependem essencialmente da forma como se define o limite de uma sucessão de variáveis

aleatórias e, no caso presente, vamos estar particularmente interessados nos limites em média quadrática e quase certo.

Um processo é **contínuo em média quadrática**, ou **contínuo em** L^2, se para qualquer $\boldsymbol{s} \in \mathbb{R}^2$,

$$\lim_{\boldsymbol{s}' \to \boldsymbol{s}} E\Big[\big(Y(\boldsymbol{s}') - Y(\boldsymbol{s})\big)^2\Big] = 0.$$

Um processo é **diferenciável em média quadrática** se para cada $\boldsymbol{s} \in \mathbb{R}^2$ existe um vector $\boldsymbol{\Delta_Y}(\boldsymbol{s})$ tal que, para qualquer escalar h e qualquer vector unitário $\boldsymbol{u}$, se tenha

$$\lim_{h \to 0} \Big[\frac{Y(\boldsymbol{s} + h\boldsymbol{u}) - Y(\boldsymbol{s}) + h\boldsymbol{u}^T \boldsymbol{\Delta_Y}(\boldsymbol{s})}{h}\Big]^2 = 0.$$

Um processo é duas vezes diferenciável em média quadrática se a sua derivada em média quadrática $Y'(\boldsymbol{s})$ é diferenciável em L^2, e assim sucessivamente para a diferenciabilidade de ordens superiores.

Se nos limitarmos ao caso isotrópico temos o seguinte resultado:

Teorema 2.1 (Bartlett, 1955). *Um processo estocástico com função de correlação $\rho(t)$ é k vezes diferenciável em média quadrática se $\rho(t)$ é $2k$ vezes diferenciável em $t = 0$.*

Passemos agora à continuidade e diferenciabilidade quase certa.

Um processo espacial é **quase certamente contínuo** se o conjunto das trajectórias que não são contínuas tem probabilidade nula, e é **quase certamente** k **vezes diferenciável** se o conjunto das trajectórias que não são k veses diferenciáveis tem probabilidade nula, (k inteiro positivo). Diz-se também que os processos têm **realizações contínuas** ou **realizações** k **vezes diferenciáveis**, respectivamente.

Os resultados sobre a diferenciabilidade quase certa foram investigados por Kent(1989) e são dados genericamente para processos

estacionários de segunda ordem em espaços de dimensão $d > 1$ não necessariamente isotrópicos. Aqui vamos apenas apresentar o caso bidimensional isotrópico.

Teorema 2.2. *Seja $r_2(t) = \rho(t) - p_2(t)$ onde $p_2(t)$ é o polinómio de grau 2 do desenvolvimento de $\rho(t)$ em série de Taylor em torno de 0. Então, se $\rho(t)$ é 2 vezes continuamente diferenciável e*

$$|r_2(t)| = O(|t|^{2+\alpha}),\ t \to 0,$$

para algum $\alpha > 0$, existe uma versão do processo espacial com realizações contínuas.

Embora qualquer uma das continuidades não implique a outra, visto as convergências não se implicarem mutuamente, ambas implicam a convergência em probabilidade e, como tal, qualquer das continuidades implica a continuidade em probabilidade do processo.

Existem, no entanto, relações entre a diferenciabilidade em média quadrática e a quase certa, se os processos forem estacionários de segunda ordem isotrópicos como é o caso nos parágrafos que se seguem.

Note-se, por exemplo, que, devido ao Teorema 2.2, se um processo espacial é quase certamente duas vezes diferenciável tem correlograma duas vezes continuamente diferenciável e, portanto, contínuo em 0. Assim sendo, o processo é contínuo em média quadrática.

Por outro lado, um processo é quase certamente contínuo se for diferenciável em média quadrática, e quase certamente diferenciável se for duas vezes diferenciável em média quadrática e assim por diante (Bartlett, 1955). Em particular se um processo é indefinidamente diferenciável em média quadrática é indefinidamente diferenciável quase certamente.

2.1.6 Famílias paramétricas de variogramas

Tal como acontece com as distribuições de variáveis aleatórias, também entre os variogramas isotrópicos válidos se distinguem classes cujos elementos podem ser descritos por uma expressão analítica comum relativamente simples, variando num pequeno número de parâmetros. Essas classes chamam-se **famílias paramétricas de variogramas**. Como no caso das distribuições, estas famílias de variogramas não são disjuntas, possuindo algumas delas, alguns modelos em comum.

Vamos passar a apresentar as famílias mais conhecidas começando por aquelas que são construídas à custa de funções de correlação válidas o que permite apresentar simultaneamente o variograma e o correspondente covariograma.

1. Família esférica

Os correlogramas que correspondem a esta família são do tipo

$$\rho(t) = \begin{cases} 1 - \frac{3}{2}\frac{t}{\phi} + \frac{1}{2}(\frac{t}{\phi})^3 & \text{se } 0 \leq t \leq \phi \\ 0 & \text{se } t > \phi \end{cases},$$

em que ϕ é um parâmetro com as mesmas unidades da distância. Recordando que

$$\gamma(t) = \tau^2 + \sigma^2(1 - \rho(t)),\ t > 0$$

e que $\gamma(0) = 0$, vem para o variograma,

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 & \text{se } t > \phi \\ \tau^2 + \sigma^2 \left[\frac{3}{2}\frac{t}{\phi} - \frac{1}{2}(\frac{t}{\phi})^3\right] & \text{se } 0 < t \leq \phi \\ 0 & \text{se } t = 0. \end{cases}$$

Tendo também em atenção que

$$C(t) = \tau^2 + \sigma^2 - \gamma(t),$$

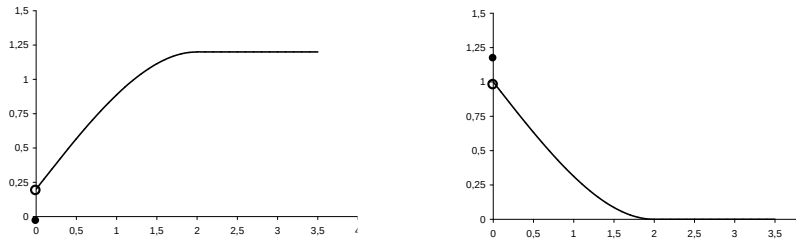
vem para o covariograma,

$$C(t) = \begin{cases} 0 & \text{se } t > \phi \\ \sigma^2 \left[1 - \frac{3}{2}\frac{t}{\phi} + \frac{1}{2}(\frac{t}{\phi})^3\right] & \text{se } 0 < t \leq \phi \\ \tau^2 + \sigma^2 & \text{se } t = 0. \end{cases}$$

Na Figura 2.1 apresenta-se, como exemplo, o gráfico do modelo esférico para $\tau^2 = 0.2, \sigma^2 = 1$ e $\phi = 2$.

Modelo esférico

$\tau^2 = 0.2$, $\sigma^2 = 1$, $\phi = 2$



Variograma Covariograma

Figura 2.1 *Modelo esférico com efeito de pepita igual a 0.2, patamar igual a 1.2 e alcance igual a 2.*

Note-se que o salto do variograma na origem é $\tau^2 = 0.2$, que o patamar é $\tau^2 + \sigma^2 = 1.2$ e que o alcance é $r = 2$.

Esta família de variogramas deve a sua popularidade ao facto de fornecer interpretações directas para os seus parâmetros, nomeadamente para ϕ, que neste caso é exactamente o alcance.

2. Família Matérn

Esta família de funções de correlação foi apresentada por Matérn em 1960 para espaços de dimensão qualquer e é também conhecida pela designação de **família Whittle-Matérn**. Ela corresponde a

um conjunto de modelos com aspectos muito variados, visto que a expressão geral da função de correlação depende de mais um parâmetro que a anterior. Esta é

$$\rho(t) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{t}{\phi}\right)^{\nu} K_{\nu}\left(\frac{t}{\phi}\right), \; t \geq 0$$

em que $K_{\nu}(.)$ é a função de Bessel modificada de ordem ν, $\nu > 0$. A **ordem** é um parâmetro de forma que, como veremos, desempenha um papel importante na regularidade das trajectórias do modelo, e ϕ é um parâmetro de escala que tem a dimensão da distância, mas que não é identificável com o alcance, que nos modelos desta família não existe.

Certas ordens particulares são frequentemente usadas visto corresponderem a expressões analíticas simples e possuem, em alguns casos, uma designação própria[6]. É o que acontece com o caso $\nu = 0.5$ conhecido pelo **modelo exponencial** que corresponde à função de correlação

$$\rho(t) = \exp(-t/\phi), \quad t \geq 0.$$

O variograma e o correlograma com efeito de pepita τ^2 e patamar parcial σ^2 são respectivamente dados por,

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2[1 - \exp(-t/\phi)] & \text{se } t > 0 \\ 0 & \text{caso contrário,} \end{cases}$$

$$C(t) = \begin{cases} \sigma^2 \exp(-t/\phi) & \text{se } t > 0 \\ \tau^2 + \sigma^2 & \text{caso contrário.} \end{cases}$$

Um outro caso em que a expressão do correlograma tem uma forma

[6]Nos casos $\nu = m + 1/2$, $m = 0, 1, \ldots$, o correlograma tem uma expressão fechada, visto que corresponde a funções característica de distribuições t-student com $\nu + 1/2$ graus de liberdade (veja por exemplo Ruppert, 2003).
O caso $\nu = 1$ é conhecido por modelo de Whittle(1954).

fechada é o de ordem $\nu = 1.5$,

$$\rho(t) = (1 + t/\phi)\exp(-t/\phi), \quad t \geq 0.$$

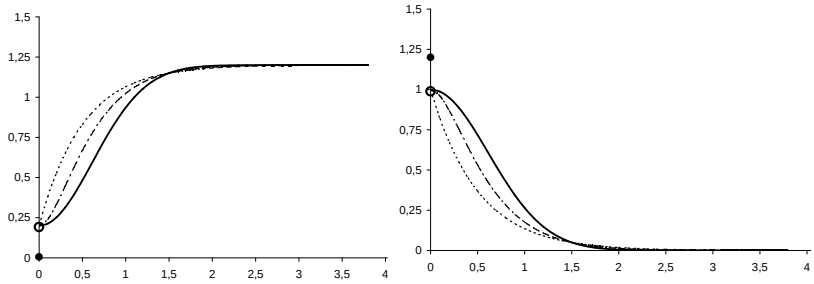
Finalmente o caso limite, obtido quando $\nu \to \infty$, é conhecido por **modelo Gaussiano** e tem correlograma dado por

$$\rho(t) = \exp\left(-\left(\frac{t}{\phi}\right)^2\right), \quad t \geq 0.$$

A Figura 2.2 mostra os semivariogramas e covariogramas dos três modelos que acabámos de referir, $\nu = 0.5,\ 1.5,\ \infty$, com o mesmo efeito de pepita, $\tau^2 = 0.2$, o mesmo patamar parcial, $\sigma^2 = 1$, e com parâmetro ϕ calculado para cada um dos modelos de forma a que o alcance efectivo seja para todos $r_0 = 1.5$.

Família Matérn: $\rho(t) = \frac{1}{2^{\nu-1}\Gamma(\nu)}\left(\frac{t}{\phi}\right)^{\nu} K_\nu(\frac{t}{\phi})$,

$\cdots\ \nu = 0.5$ $\quad$ $-\cdot-\ \nu = 1.5$ $\quad$ $—\ \nu = \infty$



Semivariograma $\qquad$ Covariograma

Figura 2.2 *Modelos da família Matérn com $\nu = 0.5,\ 1.5,\ \infty$, efeito de pepita igual a 0.2, patamar igual a 1.2 e alcance efectivo igual a 1.5.*

A forma como o parâmetro ϕ está relacionado com o alcance efectivo r_0 depende da ordem ν e, se for utilizada a definição (2.20) de

alcance efectivo, ϕ será respectivamente dado nos três modelos pelas expressões aproximadas seguintes:

$$\phi = \frac{r_0}{3}; \quad \phi = \frac{r_0}{4.75}; \quad \phi = \frac{r_0}{\sqrt{3}}.$$

De facto, parâmetros de escala para modelos com ordens diferentes não são directamente comparáveis.

No que se refere à ordem da diferenciabilidade em média quadrática do processo pode demonstrar-se (Stein, 1999) que esta é controlada pelo parâmetro ν do correlograma.

Uma vez que, de acordo com o Teorema 2.1, é a diferenciabilidade do correlograma no ponto zero que nos interessa, há que relembrar que este está definido para $\mathbb{R}$ mas que só se apresenta a definição em $\mathbb{R}^+$ por a função ser par, $\rho(-t) = \rho(t)$. Desta forma, pode verificar-se que, no caso $\nu = 0.5$ em que $\rho(t) = \exp(-|t|/\phi)$, o processo é contínuo em média quadrática mas não é diferenciável.

No caso $\nu = 1.5$, $\rho(t) = (1 + |t|/\phi)\exp(-|t|/\phi)$ é exactamente duas vezes diferenciável e, por essa razão, o processo é diferenciável em média quadrática. Note-se também que processo tem trajectórias contínuas.

Para o modelo gaussiano, o correlograma $\rho(t) = \exp(-(|t|/\phi)^2)$, é infinitamente diferenciável, o que implica trajectórias infinitamente diferenciáveis quer em média quadrática quer quase certamente [7].

De facto, estes resultados podem ser generalizados da seguinte forma: na família Matérn o maior inteiro menor ou igual a ν corresponde ao número de vezes que o processo é diferenciável em média quadrática.

[7] O processo é analítico em média quadrática o que quer dizer que se consegue prever exactamente $Y(\boldsymbol{s})$ em todo o ponto $\boldsymbol{s} \in \mathbb{R}^2$ a partir do conhecimento do processo numa vizinhança de $\boldsymbol{s}$ arbitrariamente pequena. (Veja por exemplo Banerjee, 2003)

3. Família exponencial potenciada

Esta família tem correlograma de expressão geral

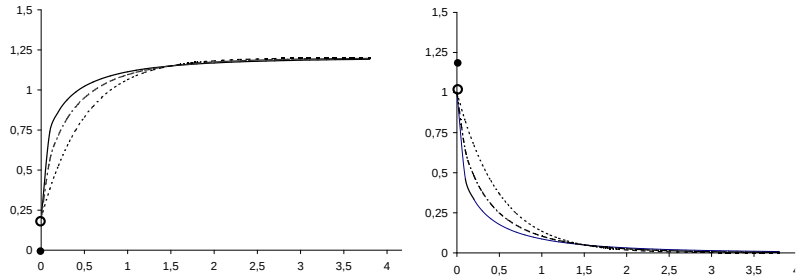
$$\rho(t) = \exp(-(t/\phi)^{\nu}), \quad t \geq 0,$$

e tal como a família Matérn tem dois parâmetros: o parâmetro de escala, $\phi > 0$ e o parâmetro de forma, ν, variando no intervalo $(0, 2]$.

Como é fácil perceber o modelo exponencial e o modelo gaussiano são também casos particulares desta família com ν respectivamente iguais a 1 e 2.

Família exponencial potenciada: $\rho(t) = \exp(-(t/\phi)^{\nu})$

$\cdots\cdots$ $\nu = 0.5$ — — $\nu = 0.7$ —— $\nu = 1$



Semivariograma Covariograma

Figura 2.3 *Modelos da família exponencial potenciada com* $\nu = 0.5, 0.7, 1$*, efeito de pepita igual a 0.2, patamar igual a 1.2 e alcance efectivo igual a 1.5.*

Na Figura 2.3 acima apresentam-se os semivariogramas e os covariogramas para a família exponencial potenciada com $\nu = 0.5,\ 0.7,\ 1$, o mesmo efeito de pepita, $\tau^2 = 0.2$, o mesmo patamar parcial, $\sigma^2 = 1$, e com parâmetro ϕ calculado para cada um dos modelos de forma a

que o alcance efectivo seja para todos $r_0 = 1.5$. O valor do parâmetro ϕ encontrado para cada um dos modelos é respectivamente igual a 0.167, 0.313 e 0.5.

Note-se que o modelo exponencial já estava presente na Figura 2.2 (linha ponteada) e foi também incluído nesta figura com a mesma linha ponteada para permitir uma comparação entre os modelos de Matérn e exponencial potenciado apresentados.

No que diz respeito à diferenciabilidade do processo verifica-se pela expressão da correlação, $\rho(t) = \exp(-(|t|/\phi)^\nu$, $0 < \nu \leq 2$, que só no caso $\nu = 2$ existe diferenciabilidade em média quadrática.

Para terminar, apresentamos mais duas famílias que possuem características que as distinguem das apresentadas até aqui.

4. Família onda

Esta família é definida pelo correlograma

$$\rho(t) = \frac{\operatorname{sen}(t/\phi)}{t/\phi}, \quad t \geq 0,$$

em que o parâmetro $\phi > 0$ tem a dimensão da distância.

Ao contrário dos exemplos anteriores este correlograma não é monotonamente decrescente e os correspondentes variograma e covariograma,

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2\left[1 - \frac{\operatorname{sen}(t/\phi)}{t/\phi}\right] & \text{se } t > 0 \\ 0 & \text{caso contrário} \end{cases}$$

$$C(t) = \begin{cases} \sigma^2\left[\frac{\operatorname{sen}(t/\phi)}{t/\phi}\right] & \text{se } t > 0 \\ \tau^2 + \sigma^2 & \text{caso contrário,} \end{cases}$$

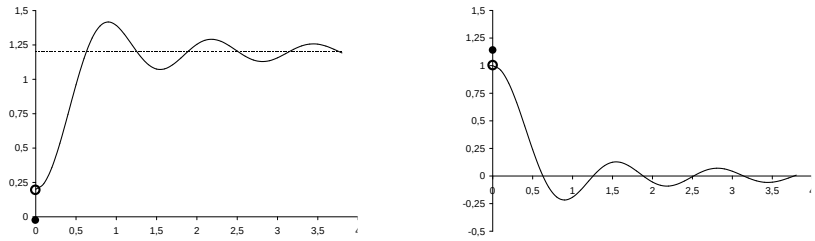
apresentam ondas que se transmitem às realizações do processo. Por essa razão este tipo de variogramas é raramente utilizado a não ser

que haja fortes suspeitas sobre o carácter ondulatório do fenómeno que se pretende modelar.

A Figura 2.4 mostra o modelo onda com efeito de pepita e patamar parcial idênticos aos anteriormente apresentados e parâmetro $\phi = 0.2$.

Modelo onda

$\tau^2 = 0.2$, $\sigma^2 = 1$, $\phi = 0.2$



Variograma Covariograma

Figura 2.4 *Modelo onda com efeito de pepita igual a 0.2, patamar igual a 1 e* $\phi = 0.2$.

5. Família potência

Trata-se de uma família de variogramas de processos intrinsecamente estacionários que não são estacionários de segunda ordem, cuja expressão geral é,

$$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t^\nu & \text{se } t > 0 \\ 0 & \text{se } t = 0 \end{cases}$$

com $\tau^2 \geq 0$, $\sigma^2 > 0$ e $\nu \in [0, 2[$.

Uma vez que se verifica,

$$\lim_{t \to +\infty} \gamma(t) = +\infty,$$

não existe covariograma.

O família com $\lambda = 1$ é conhecida por modelo linear.

A limitação para o parâmetro ν, presente também no parâmetro com o mesmo nome da família exponencial potenciada, provém da seguinte propriedade a que têm que satisfazer os variogramas válidos (Matheron, 1971):

$$\frac{2\gamma(\boldsymbol{h})}{\|\boldsymbol{h}\|} \to 0, \text{ quando } \|\boldsymbol{h}\| \to 0.$$

Existem outras famílias de variogramas conhecidas, como por exemplo a de Cauchy que é definida pelo correlograma $\rho(t) = \big(1+(t/\phi)^2\big)^{-\nu}$, $t \geq 0$, mas, para a maioria das aplicações, as famílias que apresentámos já fornecem um leque de modelos alternativos suficientemente vasto. De facto, na maioria dos casos, a dimensão do conjunto de localizações observadas não permite a identificação de modelos de variogramas muito elaborados e a família Matérn, com os seus dois parâmetros, já apresenta um conjunto de modelos com flexibilidade suficiente.

2.2 Elementos da modelação

Do ponto de vista probabilístico, o conhecimento do processo estocástico que se supõe estar na origem de um dado conjunto de observações espaciais, é dado pela especificação da lei espacial do processo, ou seja, da sua distribuição.

Isso faz-se especificando a chamada **família das distribuições conjuntas finitas** que é composta por todas as distribuições conjuntas de qualquer dimensão $n \geq 1$, correspondentes a qualquer escolha de n localizações em $\mathbb{R}^2$. Para que a especificação de um conjunto de distribuições desse género defina, de facto, a lei de um processo

estocástico é necessário que sejam válidas condições de consistência que são complicadas de verificar na prática [8].

Afastaremos este tipo de problemas restringindo as aplicações que vamos estudar às que podem ser modeladas por processos gaussianos. Nesse caso, as condições de consistência ficam asseguradas desde que se use uma função de covariância válida.

Também por essa razão, grande parte da teoria que desenvolveremos no resto do capítulo será dirigida aos processos gaussianos que passamos a definir.

Definição 2.1. *Um processo* $\{Y(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^2\}$ *é um* **processo gaussiano** *se, qualquer que seja o inteiro* $n \geq 1$ *e qualquer que seja o conjunto de localizações* $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\}$*, o correspondente vector aleatório* $(Y_1, Y_2, \ldots, Y_n)$ *tem distribuição gaussiana n-dimensional.*

Uma vez que os processos Gaussianos ficam completamente especificados pelo valor médio e pela covariância, estes processos têm a vantagem adicional de serem fortemente estacionários se forem estacionários de segunda ordem.

É evidente que muitos fenómenos não podem ser modelados por distribuições normais e, para se poder estender a utilização do modelo gaussiano a esses casos, assume-se que o modelo é válido após uma transformação dos dados.

Com dados positivos, uma classe de transformações muito usada é a de Box-Cox, (Box e Cox, 1964)

$$Z = \begin{cases} (Y^\lambda - 1)/\lambda & \lambda \neq 0, \\ \log Y & \lambda = 0. \end{cases}$$

A transformação logarítmica é talvez a mais usada na prática visto

[8] Para processos com índices em $\mathbb{R}$ estas condições são conhecidas por condições de consistência de Kolmogorov.

ser simples determinar as expressões para o valor médio e para a estrutura de covariâncias da variável transformada. Para uma exposição deste ponto referimos o leitor para o livro de Diggle e Ribeiro (2007).

2.2.1 Modelação da tendência espacial

Um dos princípios básicos da modelação espacial de dados referentes a pontos é conhecido por "**primeira lei da geoestatística**": os dados devem ser particionados num termo correspondente ao valor médio (tendência espacial) e noutro correspondente ao erro [9] ,

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \delta(\boldsymbol{s}).$$

Com esta decomposição, a variabilidade dos dados é modelada através do segundo momento dos resíduos,

$$E\Big[\big(Y(\boldsymbol{s}) - \mu(\boldsymbol{s})\big)^2\Big],$$

e a estrutura de associação espacial através do valor médio do produto,

$$E\Big[\big(Y(\boldsymbol{s}') - \mu(\boldsymbol{s}')\big)\big(Y(\boldsymbol{s}) - \mu(\boldsymbol{s})\big)\Big],$$

em vez de ser feita através dos valores médios de $\big(Y(\boldsymbol{s}) - \mu\big)^2$ e de $\big(Y(\boldsymbol{s}') - \mu\big)\big(Y(\boldsymbol{s}) - \mu\big)$, respectivamente.

A diferença entre a duas situações é medida por

$$\big(\mu - \mu(\boldsymbol{s})\big)\big(\mu - \mu(\boldsymbol{s}')\big),$$

valor que pode ser elevado se o afastamento do processo da estacionaridade de primeira ordem for grande. De facto, se o processo não tem

[9] Na notação empregue em (2.12) $\delta(\boldsymbol{s}) = W(\boldsymbol{s}) + \epsilon(\boldsymbol{s})$.

média constante, a utilização de $Y(\boldsymbol{s}) - \mu$ atribui erradamente a variação de $\mu(\boldsymbol{s})$ relativamente μ à estrutura de dependência espacial do processo.

Uma vez que a verdadeira tendência espacial é desconhecida, teremos que a estimar e, consoante a estimativa que utilizarmos, assim vamos atribuir uma maior ou menor parte da variabilidade observada nos dados à componente determinística do modelo. Esta escolha tem que ser feita cuidadosamente uma vez que, como vimos, tem repercussão na estrutura de correlação da outra componente do modelo.

Infelizmente, se não existirem observações repetidas independentes em cada localização, situação muito frequente, é impossível distinguir entre o valor médio e a realização do processo e, por conseguinte, as duas componentes do modelo não são identificáveis. Assim sendo, a escolha do modelo para a tendência tem que ser guiada por algum tipo de conhecimento externo ou por métodos estatísticos que exigem hipóteses suplementares.

Por vezes a natureza física do fenómeno em observação sugere um modelo. Noutros casos, conhecem-se variáveis explicativas dentro do domínio e pode adoptar-se para a tendência espacial um modelo de regressão linear do tipo

$$\mu(\boldsymbol{s}) = \beta_0 + \sum_{l=1}^{p} \beta_l X_l(\boldsymbol{s}),$$

em que $X_l(\boldsymbol{s})$ representa o valor da variável explicativa X_l, $(l = 1, \ldots, p)$, no local $\boldsymbol{s}$. Quando nenhuma destas situações se verifica, usam-se simplesmente polinómios de baixo grau nas coordenadas, que se designam por **superfícies de tendência**, para captar os efeitos que parecem ter uma escala de variação comparável à escala do domínio de observação, deixando a explicação de efeitos com pequena escala de variação para a outra componente do modelo.

Existem algumas técnicas mais sofisticadas para fazer esta decom-

posição da variação em variação de pequena e grande escala, como por exemplo o *median polishing algorithm* e referimos o leitor para Cressie (1993) para mais informação sobre o assunto.

Mais formalmente, a escolha entre diversas opções de modelos para a tendência espacial pode ser feita conjuntamente com a escolha do variograma usando métodos de máxima verosimilhança, como veremos mais à frente.

2.2.2 Variograma empírico

Como vimos, uma das componentes essenciais da modelação dos processos espaciais diz respeito à estrutura de associação entre as variáveis do processo.

Supondo que o processo pontual subjacente aos dados é, no mínimo, intrinsecamente estacionário e isotrópico, essa associação pode ser modelada pelo variograma omnidireccional.

A modelação do variograma pode ser comparada nos seus passos à modelação paramétrica da distribuição teórica de uma variável aleatória, suposta subjacente a um fenómeno, a partir de uma amostra de n observações: numa primeira fase utiliza-se a amostra para construir um histograma (estimador empírico da densidade) e, através dele e de algum conhecimento da natureza do fenómeno aleatório, procura-se identificar a família de distribuições que melhor representará a distribuição dos dados; seguidamente, escolhe-se para distribuição teórica da variável aleatória um dos elementos dessa família, através de um qualquer método de estimação de parâmetros. Os métodos de estimação usados são normalmente paramétricos, como é caso do método da máxima verosimilhança.

Historicamente, na modelação do variograma procede-se, até certo ponto, de forma análoga. É criado um estimador do variograma

baseado na amostra, **variograma empírico** e, com base nele, é seleccionada uma família paramétrica de variogramas, normalmente de entre as que passámos em revista, escolhendo-se ainda um ou mais modelos dessa família para eventual variograma teórico.

Neste ponto o procedimento mais clássico difere do método normal de estimação da distribuição de variáveis aleatórias que descrevemos, visto que os parâmetros do variograma teórico se estimam verificando qual dos modelos seleccionados melhor se ajusta ao variograma empírico. De facto, este método de estimação é um método de ajustamento de curvas e não um método de estimação paramétrico.

Mais recentemente, o método de estimação de variogramas usado com mais frequência está mais perto do da estimação da distribuição de variáveis aleatórias: usa-se o estimador do variograma baseado nas observações apenas como indicador da família paramétrica que deve ser seleccionada e dos valores iniciais dos correspondentes parâmetros, que são depois utilizados num método de estimação paramétrico como o da máxima verosimilhança.

Note-se que mesmo o parâmetro que distingue entre os diferentes modelos dentro de uma mesma família paramétrica, como acontece, por exemplo, com o parâmetro de forma ν da família Matérn, pode ser estimado, contudo a sua estimação não é muito consistente. Muitos autores aconselham, por isso, que sejam seleccionados, como no procedimento mais clássico, um pequeno conjunto de modelos dentro da família paramétrica e que, para esses, sejam estimados os outros parâmetros.

Evidentemente que os dois métodos têm vantagens e desvantagens, sendo que uma desvantagem do segundo é a exigência do conhecimento da lei do processo. No entanto, em muitas aplicações pode pressupor-se a distribuição gaussiana, com ou sem transformação prévia dos dados, o que possibilita uma utilização relativamente sim-

ples do método da máxima verosimilhança.

Em qualquer dos casos, o cálculo de um estimador do variograma baseado nas observações é indispensável. Passamos agora em revista os métodos mais usuais de construir esses estimadores juntamente com algumas da suas propriedades.

Já vimos que, sob a Hipótese 2.1, um processo intrinsecamente estacionário tem valor médio constante, no entanto, é frequente essa propriedade não parecer ser verificada pelos dados.

Uma vez que na estimação do variograma não é indiferente os processos serem ao não estacionários de primeira ordem, vamos supor que, quando a constância do valor médio não se verifica, se passa a estimar o variograma do processo dos resíduos, $Y(\boldsymbol{s}) - \mu(\boldsymbol{s})$, remetendo para mais tarde a forma de solucionar os problemas levantados pela necessidade de estimar a tendência espacial desconhecida $\mu(\boldsymbol{s})$.

Sob hipótese de constância do valor médio o semivariograma pode ser escrito como

$$\gamma(\|\boldsymbol{h}\|) = \frac{1}{2}E\big[\big(Y(\boldsymbol{s}) - Y(\boldsymbol{s}+\boldsymbol{h})\big)^2\big].$$

Assim, se $Y_1, Y_2, \ldots, Y_n$ são observações de um processo deste tipo, pode dizer-se que $g_{ij} = 1/2(Y_i - Y_j)^2$ é um estimador centrado do valor do semivariograma, $\gamma_{ij} = \gamma(t_{ij})$, para a distância $t_{ij} = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$.

Usando o método dos momentos, Matheron (1962) propôs o seguinte estimador para o semivariograma que é conhecido por **variograma empírico**

$$\widehat{\gamma}(t) = \frac{1}{2|N(t)|} \sum_{(\boldsymbol{s}_i, \boldsymbol{s}_j) \in N(t)} \big(Y_i - Y_j\big)^2$$

em que $N(t) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : \|\boldsymbol{s}_i - \boldsymbol{s}_j\| = t\}$ e $|N(t)|$ é o respectivo cardinal.

Note-se que, a não ser no caso em que as localizações observadas

estejam sobre uma quadrícula, é muito provável que as distâncias correspondentes a todos os pares de pontos sejam diferentes, e o variograma empírico será apenas a colecção de todos os pares $(t_{ij}, g_{ij}), j > i$.

Neste caso, o variograma empírico padece de alguns inconvenientes para poder ser um bom instrumento de estimação, nomeadamente a de os seus pontos apresentarem uma tal dispersão que se torna impossível divisar um padrão de comportamento do variograma.

Isto pode ser verificado na Figura 2.5, onde se apresenta o gráfico do variograma empírico, conhecido por **variograma nuvem**, calculado para a raiz quadrada da precipitação do mês de Novembro de 2006[10], exemplo a que já nos referimos no Capítulo 1, Figura 1.3.

Note-se que se está a usar como distância, a distância euclidiana entre os pontos de coordenadas transformadas em quilómetros. Devido às pequenas dimensões de Portugal a aproximação é aceitável.

Esta dispersão do variograma empírico para dados não quadriculados, é devida aos seguintes factos: as $n(n-1)/2$ variáveis g_{ij} são necessariamente correlacionadas visto terem sido construídas a partir de apenas n observações; o variograma é muito sensível a outliers visto que a presença de apenas uma observação discrepante contamina $n-1$ dos seus valores; no caso do modelo assumido ser gaussiano a distribuição dos g_{ij} é a de um χ^2 com 1 grau de liberdade que é altamente assimétrica.

Para podermos retirar esta enorme quantidade de "ruído" do variograma empírico é necessário proceder a algum alisamento.

Na maioria dos casos esse alisamento é conseguido, embora com algum enviesamento adicional, fazendo a média dos valores do va-

[10] Uma vez que posteriormente se pretende estimar os parâmetros do variograma pelo método da máxima verosimilhança utilizou-se uma transformação dos dados por forma a que a correspondente distribuição se aproximasse da normal. De entre as transformações de Box-Cox a raiz quadrada pareceu a mais adequada.

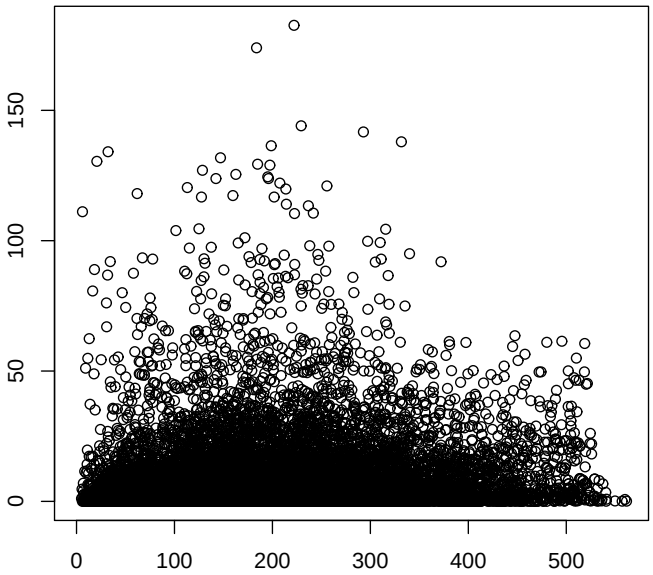


Figura 2.5 *Semivariograma nuvem da raiz quadrada da precipitação do mês de Novembro de 2006 registada em 158 estações da rede meteorológica nacional.*

riograma empírico correspondentes a distâncias incluídas em certos intervalos de distâncias contíguos.

Mais precisamente, para uma grelha de distâncias, $0 < t_1 < t_2 < \ldots < t_k$, formemos os intervalos $I_m =]t_{m-1}, t_m]$, $(m = 1, \ldots, k)$ e sejam $t'_m = (t_{m-1} + t_m)/2$ os respectivos pontos médios e $N(t'_m) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \in I_m\}$ os conjuntos de pares de pontos com distâncias em I_m. O estimador alisado do variograma constituído pelo conjunto de pontos $(t'_m, \widehat{\gamma}(t'_m))$, $m = 1, \ldots, k$, com

$$\widehat{\gamma}(t'_m) = \frac{1}{2|N(t'_m)|} \sum_{(\boldsymbol{s}_i, \boldsymbol{s}_j) \in N(t'_m)} \left[Y_i - Y_j\right]^2,$$

é designado por **variograma amostral**[11].

[11] Frequentemente, designa-se também por variograma empírico este variograma

Nesta construção é frequente escolher intervalos de igual comprimento, mas a selecção do número de intervalos, e também a escolha do último ponto, t_k, depende, evidentemente, da dimensão e estrutura das localizações amostradas, e tem problemas semelhantes aos encontrados em escolhas análogas para a construção de histogramas.

Note-se que, se por um lado os intervalos forem bastante estreitos não se está a introduzir um grande enviesamento nas estimativas intervalares, mas se, por outro, estes contiverem poucos pontos não se atingirá, com a média, uma redução significativa da variância, que é o efeito pretendido. Há, pois, que encontrar um equilíbrio que, segundo Journel e Huijbregts (1979), se consegue se a grande maioria dos intervalos tiverem a largura suficiente para conterem pelo menos 30 pares de pontos.

No variograma da Figura 2.6 procura ilustrar-se, com o exemplo da precipitação de Novembro, a forma como o número de pares de pontos em cada intervalo de distâncias, $n_m = |N(t'_m)|$, e a escolha da distância máxima influenciam o aspecto do variograma amostral.

Esta estimativa do semivariograma foi construído com 13 intervalos de igual comprimento cobrindo todas as distâncias entre pares de observações. Os centros do primeiro e último intervalos são respectivamente iguais a $t'_1 = 22\ km$ e $t'_{13} = 540\ km$.

Embora o número de pares de pontos em cada intervalo seja sempre superior a 30, esse número varia entre, $\min\{n_m\} = n_{13} = 44$ e $\max\{n_m\} = n_4 = 1977$, sendo os dos últimos cinco intervalos substancialmente mais baixos que os de todos os outros. Note-se que as estimativas correspondentes a estes últimos intervalos parecem mais variantes, o que de facto se pôde confirmar pelo respectivo desvio padrão que é naturalmente mais elevado que o das estimativas dos intervalos iniciais, não merecendo por isso tanta confiança.

---

alisado, o mesmo acontecendo ao respectivo gráfico.

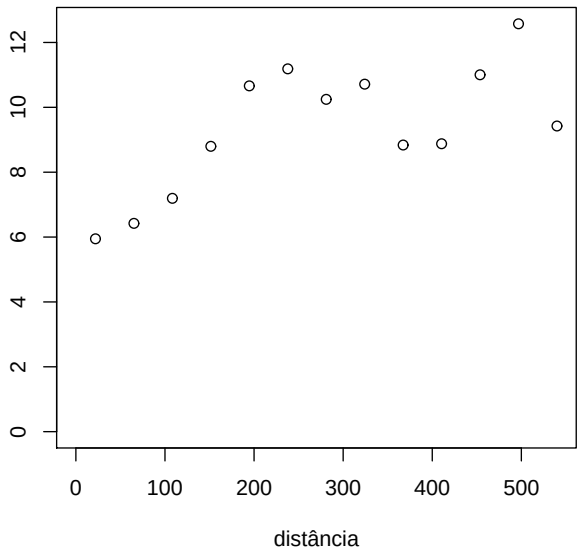


Figura 2.6 *Semivariograma amostral da raiz quadrada da precipitação com 13 intervalos e sem limitação superior das distâncias.*

Este comportamento desvantajoso é típico nos variogramas amostrais: os primeiros pontos correspondem normalmente a estimativas mais eficientes e, por isso, o semivariograma aparenta estar mais alisado para distâncias mais pequenas; os últimos pontos correspondem a estimativas de maior variância e a parte final do semivariograma tem um aspecto mais irregular.

Evidentemente que outras desvantagens que apontámos no variograma empírico se mantêm neste estimador alisado, nomeadamente a falta de robustez. Existe uma abundante literatura sobre estimadores robustos para o variograma que pode ser consultada em Cressie (1993), no entanto, vamos apenas referir 2 estimadores robustos para

a contaminação por outliers, sugeridos por Cressie e Hawkins (1980):

$$2\overline{\gamma}(t) = \left\{ \frac{1}{|N(t)|} \sum_{N(t)} |Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)|^{1/2} \right\}^4 \Big/ (0.457 + 0.494/|N(t)|)$$

$$2\widetilde{\gamma}(t) = \frac{\left[\text{med}\{|Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)|^{1/2} : (\boldsymbol{s}_i, \boldsymbol{s}_j) \in N(t)\}\right]^4}{B(t)}$$

onde $B(t)$ corrige o enviesamento e med$\{.\}$ é a mediana.

Outro tipo de estimadores alisados do variograma teórico obtêm-se através de métodos de regressão não paramétrica, como por exemplo o método de kernel ou de regressão por splines, mas a falta de independência das observações do variograma empírico cria algumas dificuldades na sua aplicação referidas, por exemplo, em Diggle e Ribeiro (2007).

Para processos estacionários de segunda ordem isotrópicos, pode definir-se analogamente o **covariograma empírico**,

$$\widehat{C}(t) = \frac{1}{|N(t)|} \sum_{(\boldsymbol{s}_i, \boldsymbol{s}_j) \in N(t)} (Y_i - \overline{Y})(Y_j - \overline{Y}),$$

onde

$$\overline{Y} = \sum_{i=1}^{n} Y_i / n$$

é um estimador do valor médio μ, e $N(t)$ e $|N(t)|$ são definidos como anteriormente.

É importante notar que para os estimadores empíricos não se tem a igualdade (2.6), ou seja, $\widehat{\gamma}(t) \neq \widehat{C}(0) - \widehat{C}(t)$.

Alguns autores apontam vantagens na utilização do variograma empírico face ao covariograma, entre elas Cressie (1993) nomeia o facto do primeiro não necessitar de estimação do valor médio do processo e de ser menos enviesado que o correspondente covariograma quando se trabalha com resíduos de uma tendência espacial.

As definições de variograma e covariograma empíricos podem ser facilmente estendidas ao caso não isotrópico usando apenas os pares de pontos com vector de separação $\boldsymbol{h}$,

$$2\widehat{\gamma}(\boldsymbol{h}) = \frac{1}{|N(\boldsymbol{h})|} \sum_{(\boldsymbol{s}_i,\boldsymbol{s}_j)\in N(\boldsymbol{h})} \big(Y_i - Y_j\big)^2, \tag{2.22}$$

em que $N(\boldsymbol{h}) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}\}$.

Caso a amostra tenha dimensão suficiente, é possível construir variogramas amostrais direccionais usando exclusivamente vectores numa dada direcção particular $\boldsymbol{h}/\|\boldsymbol{h}\|$. Veremos posteriormente como estes variogramas podem ser usados para controlo da anisotropia no contexto duma análise exploratória de dados.

Para finalizar, fazemos notar que os estimadores do semivariograma que apresentámos nem sempre satisfazem a Proposição 2.1 de serem funções semidefinidas negativas. Essa é uma das razões pela qual se ajustam aos estimadores baseados nas observações modelos de variograma válidos pertencentes às famílias paramétricas que estudámos, coisa que faremos na próxima secção.

2.3 Estimação

No que se segue estamos apenas interessados em estimar os parâmetros envolvidos nas estruturas do valor médio e da associação espacial do modelo adoptado para o processo estocástico que se supõe gerar as observações.

2.3.1 Estimação da tendência espacial

No que se refere ao processo vamos supor que ele pode ser descrito de forma genérica por

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + \delta(\boldsymbol{s}) = \beta_0 + \sum_{j=1}^{p} \beta_l X_l(\boldsymbol{s}) + \delta(\boldsymbol{s}),$$

em que $X_l,\ l = 1, \ldots, p$, são variáveis explicativas não aleatórias que podem ou não depender da localização e $\delta(\boldsymbol{s})$ é um processo de média nula e variância finita que poderá ou não ter correlação espacial.

Usaremos a notação matricial para representar o modelo referente a n localizações,

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\delta},$$

em que $\boldsymbol{Y} = \big(Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)\big)'$ é o vector de observações, $\boldsymbol{\delta} = \big(\delta(\boldsymbol{s}_1), \ldots, \delta(\boldsymbol{s}_n)\big)'$ é o vector dos resíduos, $\boldsymbol{\beta} = \big(\beta_0, \beta_1, \ldots, \beta_p\big)'$ é o vector dos coeficientes e X é a matriz $n \times (p+1)$, em que a primeira coluna tem todos os seus elementos iguais a 1 e as outras correspondem às observações nos n locais de cada uma das p covariáveis.

Se o objectivo é estimar o vector $\boldsymbol{\beta}$ eficientemente, o estimador centrado com variância mínima entre os estimadores lineares, ou seja, entre os estimadores da forma $A\boldsymbol{Y}$, é o **estimador de mínimos quadrados generalizados**. Este estimador minimiza

$$(\boldsymbol{Y} - X\boldsymbol{\beta})'\Sigma^{-1}(\boldsymbol{Y} - X\boldsymbol{\beta}),$$

e tem a expressão

$$\widehat{\boldsymbol{\beta}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\boldsymbol{Y}, \tag{2.23}$$

com matriz de variância

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = (X'\Sigma^{-1}X)^{-1}.$$

Na construção deste estimador, considera-se conhecida a matriz de covariâncias de $\boldsymbol{\delta}$, também notada por Σ, que é suposta ser invertível, acontecendo o mesmo com todas as matrizes para as quais utilizámos a notação da inversa.

Note-se que se o modelo que estamos a considerar é gaussiano, este estimador é também o estimador de máxima verosimilhança. No entanto, quando não se conhece a matriz Σ, situação em que nos encontramos no início da modelação, usa-se o **estimador de mínimos quadrados ordinários**, ou simplesmente, **estimador de mínimos quadrados**, que se obtém minimizando

$$SQR(\boldsymbol{\beta}) = \sum_{i=1}^{n} \big(Y_i - \mu(\boldsymbol{s}_i)\big)^2.$$

Este estimador é dado por

$$\widetilde{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{Y},$$

e a sua matriz de variância é dada por [12]

$$\mathrm{Var}(\widetilde{\boldsymbol{\beta}}) = (X'X)^{-1}(X'\Sigma X)(X'X)^{-1}.$$

Quando Σ é desconhecida, mas se conhece uma sua estimativa, $\widehat{\Sigma}$, pode usar-se o estimador de mínimos quadrados generalizados estimados, $\widehat{\widehat{\boldsymbol{\beta}}}$, que se obtém substituindo na expressão (2.23) a matriz de covariâncias pela sua estimativa.

Uma vez estimado $\boldsymbol{\beta}$ por algum destes métodos, a estimação dos parâmetros da estrutura de correlação do modelo prossegue então usando os resíduos

$$\widetilde{\boldsymbol{R}} = \boldsymbol{Y} - X\widetilde{\boldsymbol{\beta}}, \quad \text{ou} \quad \widehat{\widehat{\boldsymbol{R}}} = \boldsymbol{Y} - X\widehat{\widehat{\boldsymbol{\beta}}}$$

[12] A não ser no caso menos interessante dos erros não serem espacialmente correlacionados,($\Sigma = \sigma^2 I$), em que os dois estimadores são idênticos e com a mesma matriz de variância, a diferença entre as matrizes de variância, $\mathrm{Var}(\widetilde{\boldsymbol{\beta}}) - \mathrm{Var}(\widehat{\boldsymbol{\beta}})$ é semidefinida positiva o que mostra que, de facto, que o estimador generalizado é mais eficiente.

2.3.2 Estimação do variograma

Suponhamos que estamos na posse de um variograma amostral de resíduos. O problema que se vai tratar é o de escolher um modelo, de entre os fornecidos pelas diversas famílias paramétricas de variogramas, para variograma teórico do processo $\boldsymbol{\delta}(\boldsymbol{s})$.

Inicialmente, este ajustamento era feito "a olho". Procurava-se escolher um modelo paramétrico cuja aparência fosse semelhante à do variograma amostral e encontrar estimadores para os seus parâmetros: o efeito de pepita, o patamar parcial e o alcance efectivo (a partir do qual, e consoante o modelo, era estimado ϕ). Seguidamente usava-se o gráfico da curva com os parâmetros escolhidos para inspeccionar visualmente o seu ajustamento aos pontos do variograma amostral. O modelo era encontrado pelo método de tentativa e erro.

Esta técnica de estimação tem, como é referido no livro de Banerjee, Carlin e Gelfand (2004) "tanto de ciência como de arte". De facto, a inspecção do variograma amostral pode dar muitas indicações sobre o modelo e os seus parâmetros, mas a qualidade do ajustamento é algo que precisa ser medido de forma mais precisa que por simples inspecção visual.

Tomemos, para começar, o efeito de pepita. Este precisa ser estimado por extrapolação do variograma amostral no ponto zero, visto que na maioria dos casos não existem observações repetidas para ser estimado o erro de medição [13]. Ora, pode acontecer que modelos de famílias diferentes pareçam ajustar-se igualmente bem aos pontos do variograma amostral e conduzam a estimativas muito diferentes do efeito de pepita, que é um parâmetro essencial, como vamos ver posteriormente, na predição espacial.

[13] Note-se que o efeito de pepita pode também configurar, como já referimos, alguma covariância de escala inferior à menor distância entre pontos.

Da mesma forma, existem dificuldades na estimação visual do valor do patamar. É muito vulgar os valores do variograma amostral para as distâncias maiores apresentarem grandes flutuações que discordam bastante dos valores para distâncias médias. Isso deve-se ao facto destas estimativas serem calculadas com um número de pares de pontos que é, frequentemente, muito inferior ao número utilizado no cálculo das estimativas das outras distâncias, o que lhes traz uma variância superior. Por essa razão os últimos pontos não são muitas vezes tomados em consideração no ajustamento do variograma ou, equivalentemente, toma-se o limite superior do último intervalo, t_k, inferior à distância máxima entre observações. No entanto, não existe nenhuma regra objectiva para dizer qual a distância máxima que se deve usar, ou quais os pontos que se devem excluir do ajustamento. Recorde-se que este problema foi abordado no exemplo da Figura 2.6.

Finalmente, se for feito o semivariograma dos resíduos para diferentes modelos de tendência espacial, verificar-se-á que o seu formato pode variar bastante. Quanto maior é a parte da variância explicada pela tendência espacial menor tenderá a ser a parte explicada pelo erro o que fará com que o patamar tenda a ser mais baixo e a ser atingido mais depressa. O efeito de pepita contudo tenderá a permanecer inalterado.

Para ilustrar este aspecto apresenta-se na Figura 2.7 o semivariograma da raiz quadrada da precipitação, já apresentado na Figura 2.6, e o correspondente variograma dos resíduos relativos a uma superfície de tendência de primeiro grau estimada pelo método de mínimos quadrados ordinários,

$$\widetilde{R}_i = Y_i - (\widetilde{\beta}_0 + \widetilde{\beta}_1 \lambda_i + \widetilde{\beta}_2 \varphi_i),$$

em que, como habitualmente, λ representa a longitude e φ a latitude.

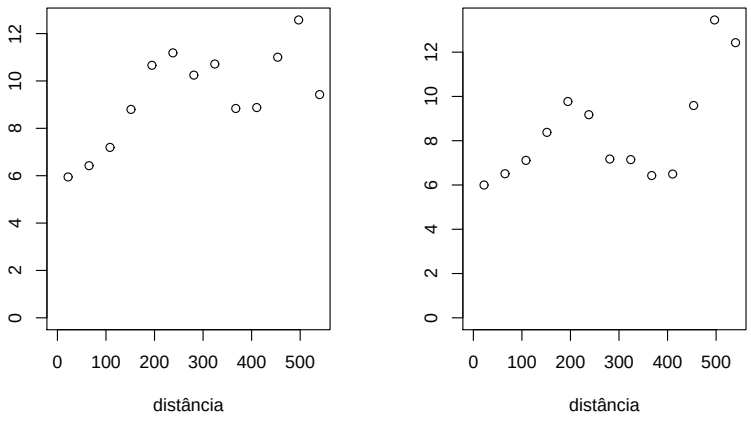


Figura 2.7 *Variogramas amostrais da raiz quadrada da precipitação (à esquerda) e dos resíduos relativos a uma superfície de tendência do primeiro grau (à direita).*

Note-se que o efeito de pepita parece estar nos dois casos em torno do valor 6, mas que o patamar, que no primeiro variograma amostral parece ser atingido por volta dos 200 km e rondar o valor 10, decresceu no segundo para um valor difícil de identificar, possivelmente em torno de 8, atingido a uma distância bastante mais curta.

Aliás, se excluirmos da análise os últimos pontos de ambos os variogramas amostrais cuja estimação não é, como vimos, tão eficiente, o variograma referente aos dados fica com um aspecto acentuadamente crescente o que pode levar a pensar que o processo não tem valor médio constante, podendo uma boa parte da correlação espacial ser explicada por uma superfície de tendência. O segundo variograma parece confirmar esta hipótese visto que os valores obtidos são muito irregulares e sem definição de um patamar claro.

Na maioria dos casos, os variogramas amostrais apresentam no início um crescimento rápido, atingindo depois um nível mais ou menos constante, o patamar. Por vezes, esse nível parece ser atingido a

partir de distâncias relativamente curtas. Nesse caso, pode estar-se perante um modelo do tipo $Y_i = \mu(\boldsymbol{s_i}) + Z_i$ em que os Z_i são variáveis aleatórias independentes, ou seja, $\boldsymbol{\delta}(\boldsymbol{s})$ não apresenta correlação espacial.

Visualmente, é difícil identificar uma situação deste tipo, no entanto, existem outras formas de testar a independência dos resíduos. Uma delas assenta na construção do chamado **invólucro do variograma amostral** pelo método Monte Carlo. Trata-se de simular m variogramas amostrais com base nas observações, como se estas fossem independentes. Cada simulação corresponde a uma permutação dos resíduos entre as localizações. Seguidamente, para cada abcissa, calcula-se o máximo e o mínimo dos valores obtidos nos m variogramas amostrais simulados. O invólucro é o gráfico desses valores para todas as abcissas.

Uma vez que sob a hipótese de independência as simulações são igualmente prováveis a probabilidade de um ponto do variograma amostral estar fora do invólucro é $\frac{1}{s+1}$.

É natural que os valores do variograma correspondentes a distâncias mais elevadas estejam dentro do invólucro visto que a correlação entre pontos distantes deve ser pequena, mas se também não existem pontos exteriores ao invólucro para distâncias pequenas, o mais certo é que a correlação espacial não seja significativa.

Os últimos comentários sobre o exemplo que estamos a estudar mostram a necessidade de inspeccionar a hipótese de independência através dos invólucros de ambos os variogramas.

A Figura 2.8, mostra os invólucros pretendidos e também o correspondente variograma amostral dos resíduos de uma segunda superfície de tendência desta feita quadrática

$$\beta_0 + \beta_1\lambda + \beta_2\varphi + \beta_3\lambda^2 + \beta_4\varphi^2,$$

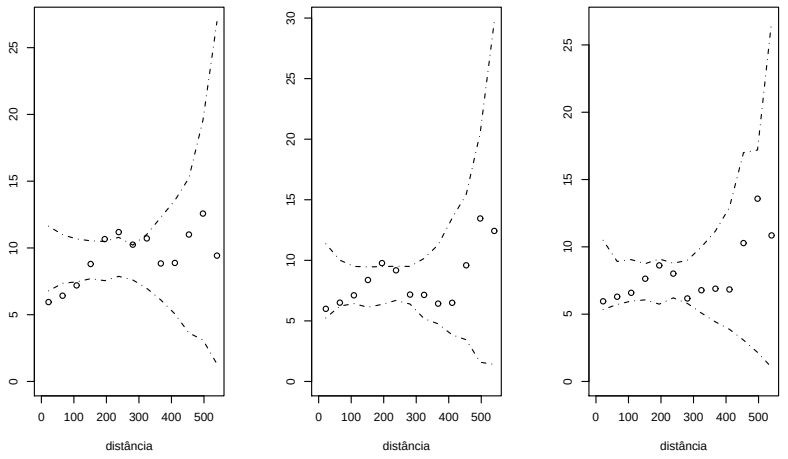


Figura 2.8 *Invólucros dos variogramas amostrais: raiz quadrada da precipitação (à esquerda), resíduos relativos a uma superfície de tendência do primeiro grau (ao centro) e resíduos relativos a uma superfície de tendência do segundo grau (à direita).*

construídos à custa de 999 simulações.

Como se pode verificar, o invólucro do variograma dos dados não contém os seus pontos iniciais, sugerindo que os dados são espacialmente correlacionados até uma distância média. Os dois invólucros seguintes sugerem que, depois de retirada uma superfície de tendência, os resíduos se podem considerar como aproximadamente independentes, não existindo uma clara diferença entre o efeito do modelo linear e o do quadrático.

De facto, embora no modelo quadrático todas as variáveis sejam significativas, ao contrário do modelo linear onde apenas a latitude o é, a proporção de variância explicada ajustada aumenta apenas de 5% no modelo mais simples para o mais complexo. Não existindo uma vantagem inequívoca entre a tendência linear e a quadrática, o princípio da parcimónia leva-nos a preferir o primeiro modelo.

Um outro problema que pode ser abordado através do estudo dos variograma amostrais é o da anisotropia.

Como dissemos anteriormente podem construir-se variogramas empíricos direccionais, mas em amostras não quadriculadas é difícil encontrar pares de pontos com vectores de separação exactamente com a mesma direcção. Desta forma, para fazer o variograma amostral numa dada direcção de ângulo η, escolhem-se todos os pares de pontos cujo vector de separação tenha direcção idêntica à direcção pretendida ou que com ela faça um pequeno ângulo ϵ, ou seja, todos os pares de pontos cujo vector de separação tenha direcção no intervalo $]\eta - \epsilon, \eta + \epsilon]$.

É, por exemplo, vulgar considerar as seguintes quatro direcções, de ângulos $\eta_i = 0^o, 45^o, 90^o, 135^o$, medidos relativamente à parte positiva do eixo dos yy no sentido directo, e construir os variogramas segundo cada uma delas, com pares de pontos com vector de separação de direcção dentro dos ângulos $\eta_i \pm 22.5^o$. Caso o aspecto dos variogramas amostrais seja marcadamente diferente para cada uma destas direcções é provável a presença de anisotropia.

A anisotropia que é mais fácil de detectar através dos variogramas direccionais é a anisotropia geométrica, pois fica notório que o patamar se atinge a distâncias diferentes em cada direcção.

É, no entanto, necessário ter alguma prudência na decisão de considerar o fenómeno como anisotrópico com base em alguma variação dos variogramas direccionais pelas seguintes razões:
– os variogramas direccionais são construídos com um número de pares de pontos bastante inferior ao do variograma omnidireccional e, como tal, quando o número de observações não é muito numeroso, as suas estimativas, nomeadamente as referentes a distâncias maiores, são muito pouco fiáveis;
– estudos de simulação mostram que, mesmo para um processo isotrópico,

pode existir uma variação apreciável entre os variogramas direccionais.

Estas desvantagens levam-nos a sublinhar de novo que apenas se devem utilizar os variogramas numa fase de exploração de dados para sugerir boas hipóteses sobre o comportamento mais provável do fenómeno.

Regressando ao exemplo da precipitação que temos vindo a seguir, apresentamos na Figura 2.9 os variogramas amostrais segundo as direcções mais usuais que referimos e também o variograma amostral omnidireccional.

Propositadamente terminámos os variogramas à distância 450 km, inferior á distância máxima verificada na amostra. A partir desta distância os variogramas tinham um comportamento demasiado instável para ter significado.

Como pode verificar-se, com excepção do variograma correspondente à direcção Este-Oeste, todos os variogramas têm aproximadamente o mesmo aspecto, estando o da direcção norte-sul particularmente perto do variograma omnidireccional.

Note-se que no modelo linear para a superfície de tendência que construímos, a latitude (N-S) é uma direcção significativa, enquanto que a longitude (E-W) não parece influenciar a precipitação. De facto, o variograma correspondente a esta última direcção parece estar associado a um conjunto de variáveis aleatórias independentes sendo por isso a correlação espacial somente determinada pela direcção norte-sul [14].

Pelas razões que apontámos, ficou bem patente a importância do variograma amostral como ferramenta de diagnóstico, mas ficaram igualmente claras as limitações que os métodos de estimação baseados

[14] Poder-se-ia esquecer a coordenada longitude e modelar o fenómeno como um processo evoluindo numa única direcção.

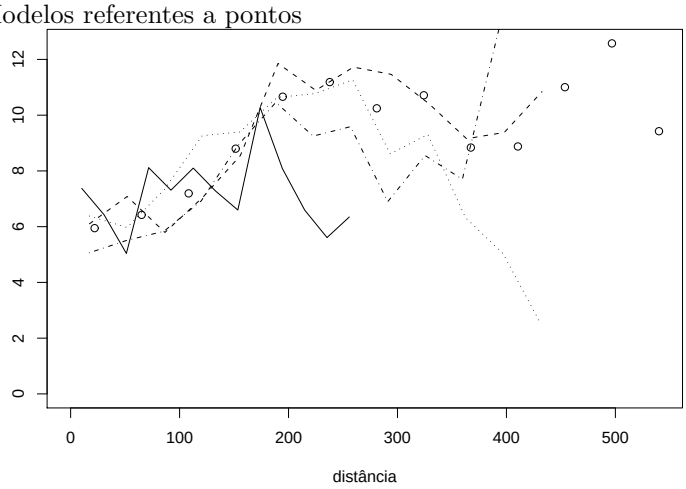


Figura 2.9 *Variograma amostral omnidireccional e variogramas amostrais nas direcções:* $\cdots\cdot$ 0°, - - - 45°, $-$ 90° *e* $-\cdot-\cdot$ 135°.

nestes variogramas podem apresentar.

Nos próximos parágrafos vamos passar em revista alguns métodos que usam este tipo de abordagem e, seguidamente, apresentaremos o método da máxima verosimilhança aplicado a modelos gaussianos que prescinde destes estimadores.

Método de mínimos quadrados

Neste parágrafo vamos supor que já está identificada a tendência espacial e que o variograma amostral dos resíduos está disponível.

Simplificaremos um pouco a notação usada até aqui: os pontos do variograma amostral serão (d_m, g_m), sendo n_m o número de pares de observações usados no cálculo de g_m, $m = 1, \ldots, k$.

O objectivo é estimar o vector de parâmetros, $\boldsymbol{\theta}$, de um modelo de variograma teórico que será notado por $\gamma(d, \boldsymbol{\theta})$, $d \geq 0$, através de um método de ajustamento da curva γ ao variograma amostral.

O método de ajustamento mais conhecido é o de mínimos quadra-

dos: os valores de $\boldsymbol{\theta}$ obtêm-se minimizando

$$\sum_{m=1}^{k} \big(g_m - \gamma(d_m;\boldsymbol{\theta})\big)^2.$$

Por vezes usa-se um outro estimador de mínimos quadrados, em que o termo em m é pesado por n_m com o intuito de controlar as diferenças existentes entre o número de pares de observações que contribuem para o cálculo de cada g_m. Neste caso minimiza-se

$$\sum_{m=1}^{k} n_k\big(g_m - \gamma(d_m;\boldsymbol{\theta})\big)^2$$

obtendo-se um estimador que não deve estar muito afastado do estimador do variograma teórico, que fosse calculado pelo mesmo método mas usando o variograma empírico.

A Figura 2.10 mostra o variograma amostral e os modelos exponencial e gaussiano para o variograma teórico obtidos pelos dois métodos de estimação de mínimos quadrados que acabámos de apresentar.
No lado esquerdo figuram os modelos estimados pelo método de mínimos quadrados ordinários e do lado direito os estimados pelo método de mínimos quadrados ponderados por n_m.

A figura sugere que qualquer dos modelos se ajusta bem aos dados, embora produzam curvas com algumas diferenças mais aparentes para pequenas distâncias. Essas diferenças também são visíveis nos estimadores de $\boldsymbol{\theta}$ que se encontram na Tabela 2.1 juntamente com o mínimo da soma dos quadrados alcançado em cada caso.

Note-se que o patamar é estimado por valores em torno de 10, tendo o efeito de pepita pesos diferentes nesse total. Os valores mais altos para este parâmetro são sempre os do modelo gaussiano que é o preferido pelos dois métodos.

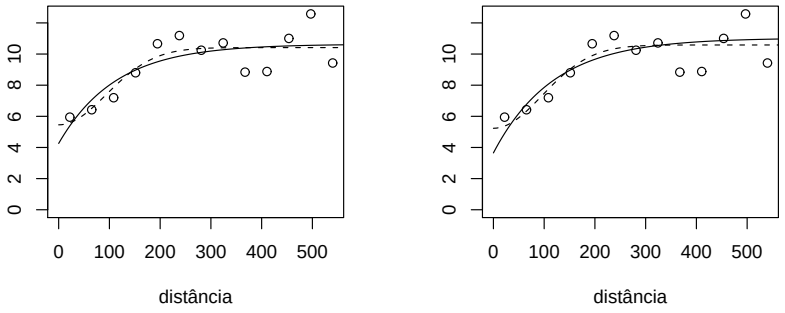


Figura 2.10 *Variograma amostral da raiz quadrada da precipitação e modelos exponencial (linha contínua) e gaussiano (linha tracejada) estimados pelo método de mínimos quadrados ordinários (à esquerda) e de mínimos quadrados ponderados pelo número de pares de observações em cada intervalo (à direita).*

Método de mínimos quadrados ordinários				
Modelo	τ^2	σ^2	ϕ	MSQ
Exponencial	4.26	6.38	112.86	15.65
Gaussiano	5.45	4.96	132.95	13.23

Método de mínimos quadrados ponderados				
Modelo	τ^2	σ^2	ϕ	MSQ
Exponencial	3.65	7.38	118.60	10102.07
Gaussiano	5.23	5.35	135.85	6671.01

Tabela 2.1 *Estimativas dos parâmetros para o modelo exponencial e gaussiano através de dois métodos de mínimos quadrados.*

Método de mínimos quadrados ponderados

Muitos outros métodos de mínimos quadrados ponderados têm sido propostos, mas vamos apenas referir brevemente o mais conhecido, proposto por Cressie(1985), que pretende levar em consideração que a variância de cada ordenada do variograma amostral deve depender

não só de n_m como também do valor correspondente do variograma teórico.

A expressão a minimizar passa a ser a seguinte:

$$\sum_{m=1}^{k} n_k \left(\frac{g_m - \gamma(d_m; \boldsymbol{\theta})}{\gamma(d_m; \boldsymbol{\theta})} \right)^2 .$$

Este estimador tem recebido alguma contestação provocada, em particular, pelo facto da equação de estimação ser enviesada e por essa razão alguns autores não o recomendam.

Considerações gerais sobre os métodos de ajustamento de curvas

Outros tipos de métodos para estimação dos parâmetros da correlação espacial através do ajustamento de curvas ao variograma amostral têm sido sugeridos, mas o facto das ordenadas do variograma não corresponderem a observações independentes cria problemas na aplicação de métodos de regressão não linear. Por essa razão muitos autores preferem fazer hipóteses adicionais sobre a distribuição do modelo subjacente ao fenómeno e usar métodos de máxima verosimilhança aproveitando o conhecimento adquirido na análise exploratória do variograma amostral.

Deve também chamar-se a atenção para o facto de não se estarem a usar para estimar os parâmetros do variograma os resíduos verdadeiros,

$$\boldsymbol{R} = \boldsymbol{Y} - X\boldsymbol{\beta},$$

mas sim uma aproximação dos resíduos,

$$\widetilde{\boldsymbol{R}} = \boldsymbol{Y} - X\widetilde{\boldsymbol{\beta}}.$$

Pode verificar-se que os variogramas calculados com base nos dois tipos de resíduos não estão normalmente muito afastados, embora o

calculado com base em $\widetilde{\boldsymbol{R}}$ esteja geralmente abaixo do obtido com os resíduos verdadeiros. Naturalmente que estas diferenças tendem a esbater-se à medida que o número de observações aumenta e as estimativas dos parâmetros ficam mais perto dos valores verdadeiros.

O tema da convergência dos estimadores do variograma está fora do alcance deste curso, chama-se apenas a atenção que o comportamento assintótico dos estimadores pode ser analisado em dois contextos diferentes: aumento da densidade de pontos observados dentro de um domínio fixo (*infill asymptotics*); aumento do domínio mantendo a densidade de pontos de observação (*increasing-domain asymptotics*).

Sobre o tema recomenda-se a leitura do artigo Zhang e Zimmerman (2005) e das referências nele contidas.

Método da máxima verosimilhança

Se for adoptado um modelo gaussiano para o processo subjacente aos dados, ou para uma transformação destes, pode utilizar-se o método da máxima verosimilhança para estimar todos os parâmetros do modelo sem recorrer a nenhum tipo de variograma estimado.

Suponhamos então que $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ tem distribuição gaussiana

$$\boldsymbol{Y} \sim \mathcal{N}\big(X\boldsymbol{\beta}, \Sigma(\boldsymbol{\theta})\big) \tag{2.24}$$

em que X é a matriz de observações de $p\,(< n)$ covariáveis, $\boldsymbol{\beta}$ é o correspondente vector de parâmetros da regressão linear e que $\Sigma(\boldsymbol{\theta})$ é a matriz de correlação que escreveremos sob a forma

$$\Sigma(\boldsymbol{\theta}) = \tau^2 I + \sigma^2 H(\boldsymbol{\phi}),$$

em que a matriz H depende de um parâmetro escalar ou vectorial $\boldsymbol{\phi}$.

Estimativas de todos os parâmetros do modelo obtêm-se maximi-

zando o logaritmo da verosimilhança dado por

$$L(\boldsymbol{\beta}, \tau^2, \sigma^2, \boldsymbol{\phi}) = -\frac{1}{2}\big\{n\log(2\pi) + \log\big(|\tau^2 I + \sigma^2 H(\boldsymbol{\phi})|\big) \qquad (2.25)$$
$$+(\boldsymbol{y} - X\boldsymbol{\beta})'\big(\tau^2 I + \sigma^2 H(\boldsymbol{\phi})\big)^{-1}(\boldsymbol{y} - X\boldsymbol{\beta})\big\}$$

Existem várias técnicas de maximização que podem ser aplicadas neste caso, no entanto, a que é mais comum baseia-se nas ideias de Mardia e Marshall (1984) e Kitadinis e Lane (1985) de aplicar o algoritmo de Gauss-Newton para minimizar o simétrico do logaritmo da verosimilhança para todos os parâmetros simultaneamente.

Este método conduz ao algoritmo que se descreve de seguida.

Reparametrize-se primeiro a matriz de covariâncias para $\varrho^2 = \tau^2/\sigma^2$ e escreva-se $V = \varrho^2 I + H(\boldsymbol{\phi})$. Dada V, a função de verosimilhança é maximizada em

$$\widehat{\boldsymbol{\beta}}(V) = (X'V^{-1}X)^{-1}X'V^{-1}\boldsymbol{y}, \qquad (2.26)$$

estimativa dos mínimos quadrados generalizados de $\boldsymbol{\beta}$, e

$$\widehat{\sigma}^2(V) = n^{-1}\big(\boldsymbol{y} - X\widehat{\boldsymbol{\beta}}(V)\big)'V^{-1}\big(\boldsymbol{y} - X\widehat{\boldsymbol{\beta}}(V)\big). \qquad (2.27)$$

Substituindo estas expressões em (2.26) obtém-se o logaritmo da verosimilhança reduzida,

$$L_0(\varrho^2, \boldsymbol{\phi}) = -\frac{1}{2}\big\{n\log(2\pi) + n\log\widehat{\sigma}^2(V) + \log|V| + n\big\},$$

que é maximizada numericamente para ϱ^2 e $\boldsymbol{\phi}$.

Seguidamente substituem-se estas estimativas em (2.26) e (2.27).

Os detalhes desta técnica podem variar com os modelos paramétricos da covariância teórica que estão a ser utilizados e alguns autores aconselham a que, na prática, se reduza a escolha do modelo paramétrico a uma família relativamente vasta como a de Matérn, mas com o parâmetro de forma ν fixado em alguns valores, ao invés de

procurar também estimar este parâmetro. Aliás, como já adiantámos anteriormente, ν está relacionado com o alisamento das trajectórias do processo, e algum conhecimento prévio do fenómeno em estudo pode dar indicações sobre os valores de ν mais prováveis. Desta forma evita-se o problema causado pela forte correlação que os estimadores de ν e ϕ tendem a mostrar, o que pode causar problemas na maximização de L_0.

Finalizamos este tema chamando a atenção para que, embora os estimadores de máxima verosimilhança sejam, como se sabe, assintoticamente normais, centrados e de variância mínima sob condições de regularidade não muito restritivas, alguns autores têm referido que os estimadores de máxima verosimilhança dos parâmetros da matriz de correlação podem ser bastante enviesadas para amostras pequenas. Este problema que pode ser minorado com o uso do método da máxima verosimilhança restrita que apresentamos de seguida.

Acerca do tema da convergência, que como dissemos não vamos abordar, recomenda-se o trabalho de Mardia e Marshall (1984) que já citámos a propósito destes estimadores.

Método da máxima verosimilhança restrita

O método da máxima verosimilhança restrita (REML) foi desenvolvido por Patterson e Thompson (1971, 1974) e pode ser aplicado com vantagem no contexto da estimação dos parâmetros do modelo gaussiano espacial. Resume-se seguidamente a ideia geral do método na sua forma mais simples.

Sob a hipótese do valor médio ser efectivamente como em (2.24) pode encontrar-se uma transformação linear de $\boldsymbol{Y}$ de tal forma que a distribuição dos dados transformados não dependa de $\boldsymbol{\beta}$. Designemos a matriz $(n-p) \times n$ dessa transformação por A' e seja $\boldsymbol{W} = A'\boldsymbol{Y}$ o

vector transformado[15].

Como se sabe, $\boldsymbol{W}$ continuará a ter uma distribuição normal mas, desta feita, só dependente de $\boldsymbol{\theta}$. Pode agora usar-se a verosimilhança escrita com base nos dados transformados $L_{\boldsymbol{W}}(\boldsymbol{\theta})$ para encontrar o estimador REML para $\boldsymbol{\theta}$. De facto, as estimativas REML de $\boldsymbol{\theta}$ são independentes da transformação escolhida desde que tenha a propriedade requerida, e usando a matriz A' que satisfaz

$$AA' = I - X(X'X)^{-1}X' \;\text{ e }\; A'A = I,$$

e que está nas condições necessárias, encontra-se uma expressão para $L_{\boldsymbol{W}}(\boldsymbol{\theta})$ que não faz intervir A' (nem $\boldsymbol{\beta}$) e que pode ser maximizada para $\boldsymbol{\theta}$.

Para estimativas de $\boldsymbol{\beta}$ usam-se as dos mínimos quadrados generalizados estimados com base na matriz de covariâncias com os parâmetros substituídos pelas suas estimativas REML ($\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_{\text{REML}}$) que, como se sabe, coincidem com as estimativas de máxima verosimilhança de $\boldsymbol{\beta}$ para $\Sigma(\widehat{\boldsymbol{\theta}}_{\text{REML}})$.

Existem melhoramentos deste método básico, mas vários autores apontam que todos eles produzem estimadores de $\boldsymbol{\theta}$ que, sendo menos enviesados que os estimadores de máxima verosimilhança, são mais sensíveis que estes ao modelo escolhido para a tendência espacial.

A Figura 2.11 refere-se ao exemplo da precipitação de Novembro de 2006 que temos vindo a seguir, e nela podem comparar-se os variogramas teóricos estimados pelo método da máxima verosimilhança e pelo da máxima verosimilhança restrita para os dois modelos da família de Matérn que já estudámos anteriormente: o modelo exponencial e o de Gauss.

Como se pode verificar os dois métodos produzem curvas distintas

[15] Note que se tem $A'X = 0$.

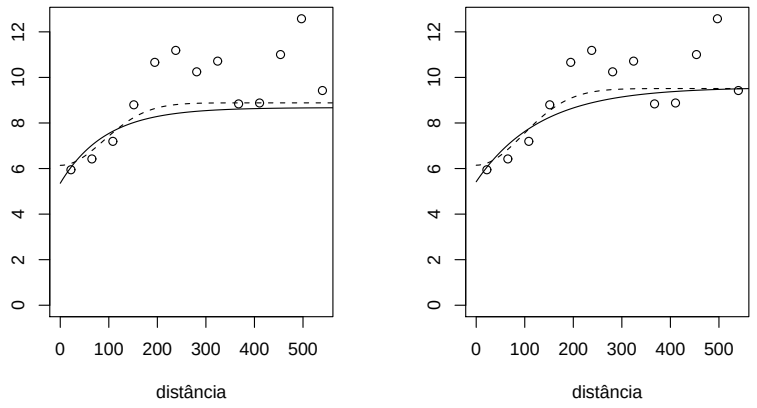


Figura 2.11 *Variograma amostral e modelos exponencial (linha contínua) e gaussiano (linha tracejada) estimados pelo método da máxima verosimilhança (à esquerda), e pela máxima verosimilhança restrita (à direita).*

para os modelos de variograma considerados. Essas diferenças entre curvas parecem ser mais acentuadas quando se utiliza o método da máxima verosimilhança restrita, e são também mais acentuadas que as encontradas quando se utilizou os métodos de mínimos quadrados. Por outro lado, também parecem ser mais acentuadas que anteriormente as diferenças entre variogramas do mesmo modelo estimados pelos dois métodos.

Na Tabela 2.2 mostram-se as estimativas dos parâmetros e também o máximo da verosimilhança obtido para cada um dos modelos e métodos. Para qualquer dos modelos o método da máxima verosimilhança restrita parece produzir melhores estimadores e de entre os modelos parece ser o gaussiano o que se adapta melhor aos dados. Note-se, no entanto, que as superfícies geradas por um modelo deste género são infinitamente diferenciáveis o que vem mais uma vez apontar no sentido de o processo poder ser bem modelado por uma superfície de tendência associada a um ruído branco.

Método da máxima verosimilhança					
Modelo	μ	τ^2	σ^2	ϕ	Max. ver.
Exponencial	13.30	5.36	3.32	93.25	-378.4
Gaussiano	13.28	6.14	2.74	125.81	-378.1

Método da máxima verosimilhança restrita					
Modelo	μ	τ^2	σ^2	ϕ	Max. ver.
Exponencial	13.22	5.42	4.15	131.74	-375.1
Gaussiano	13.52	6.15	3.37	135.55	-374.8

Tabela 2.2 *Estimativas dos parâmetros para o modelo exponencial e gaussiano através de dois métodos de máxima verosimilhança.*

2.4 Predição espacial clássica

Nesta secção vamos apresentar a abordagem clássica à predição espacial que é baseada na minimização do **erro quadrático médio**. Este conjunto de métodos de predição é conhecido por **kriging**, nome atribuído por Matheron em honra do engenheiro sul-africano Daniel G. Krige, cujo trabalho "A statistical approach to some basic mine valuation problems on the Witwatersrand"(1951), inspirou a abordagem geral do problema.
Formalmente, a questão é a de predizer, com base em n observações $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))' = (Y_1, \ldots, Y_n)'$ de um campo aleatório, uma realização de uma variável aleatória $T = T\big(Y(\boldsymbol{s})\big)$, com $\boldsymbol{s}$ variando no domínio $\mathcal{D}$. No caso mais simples, a variável T pode ser $Y(\boldsymbol{s}_0)$, ou uma sua versão sem ruído, mas também é usual T ser a variável $\int_B Y(\boldsymbol{s})d\boldsymbol{s}/|B|$ em que $|B|$ é a área de um qualquer conjunto $B \subset \mathcal{D}$.

Passemos agora à abordagem geral do problema.

Seja $t(\boldsymbol{Y})$ a função das observações que se pretende encontrar para predizer T, ou seja, o **preditor**. Neste conjunto de métodos, o melhor

preditor determina-se minimizando a função de perda erro quadrático médio, sendo assim dado pela função t que minimiza

$$E[\big(T - t(Y)\big)^2].$$

É bem conhecido da teoria bayesiana da decisão que este preditor pode também obter-se minimizando o erro quadrático médio condicionado pelas observações, $E\big[\big(T - t(\boldsymbol{Y})\big)^2|\boldsymbol{Y}\big]$, e, para isso, vamos desenvolver esta expressão adicionando e subtraindo $E[T|\boldsymbol{Y}]$.

Levando em conta que o termo cruzado é nulo vem,

$$E\big[\big(T - t(\boldsymbol{Y})\big)^2|\boldsymbol{Y}\big] = E\big[\big(T - E[T|\boldsymbol{Y}]\big)^2|\boldsymbol{Y}\big] + E\big[\big(E[T|\boldsymbol{Y}] - t(Y)\big)^2|\boldsymbol{Y}\big].$$

Visto que a primeira parcela do segundo membro desta igualdade não depende de t, podemos concluir que a função que minimiza o erro quadrático médio é

$$t(\boldsymbol{Y}) = E[T|\boldsymbol{Y}].$$

O mínimo do erro quadrático médio é designado por **variância de predição** ou **variância de kriging** e é notado por $\mathrm{Var}(T|\boldsymbol{Y})$,

$$\min E\big[\big(T - t(\boldsymbol{Y})\big)^2|\boldsymbol{Y}\big] = E[(T - E[T|\boldsymbol{Y}])^2|\boldsymbol{Y}] = \mathrm{Var}(T|\boldsymbol{Y}).$$

Restrinjamos o problema da predição ao de $T = Y(\boldsymbol{s}_0) = Y_0$ com $\boldsymbol{s}_0 \in \mathcal{D}$. Evidentemente que o estimador $E[Y_0|\boldsymbol{Y}]$ poderá não ser linear em $\boldsymbol{Y}$ e, por isso, será interessante saber qual será o melhor preditor linear de acordo com a mesma função de perda.

Um preditor linear de Y_0 baseado em $\boldsymbol{Y}$ terá a expressão

$$\sum_{i=1}^{n} l_i Y_i + k_0, \tag{2.28}$$

e o melhor preditor linear deve minimizar

$$E\Big[\big(Y(\boldsymbol{s}_0) - (\sum_i l_i Y(\boldsymbol{s}_i) + k_0)\big)^2\Big] \tag{2.29}$$

relativamente a k_0 e aos l_i.

Note-se que se o processo é estacionário de primeira ordem com valor médio μ conhecido,

$$E\Big[\big(Y(\boldsymbol{s}_0) - (\sum_i l_i Y(\boldsymbol{s}_i) + k_0)\big)^2\Big] = \mathrm{Var}\Big(Y_0 - \sum_i l_i Y_i\Big) + \big(\mu - \sum_i l_i \mu - k_0\big)^2.$$

Uma vez que o primeiro termo do segundo membro não depende de k_0, os coeficientes do melhor preditor terão que satisfazer a igualdade

$$k_0 = \mu(1 - \sum_i l_i).$$

Se se conhecer o vector $\boldsymbol{C_0} = \big(\mathrm{Cov}(Y_0, Y_1), \ldots, \mathrm{Cov}(Y_0, Y_n)\big)'$ e ainda a matriz Σ de entradas $\mathrm{Cov}(Y_i, Y_j)$, $i, j = 1, \ldots, n$, pode demonstrar-se que o melhor preditor linear é dado por

$$\widehat{Y}_0 = \boldsymbol{C}_0' \Sigma^{-1} (\boldsymbol{Y} - \mu \mathbf{1}) + \mu \tag{2.30}$$

com variância de predição

$$\mathrm{Cov}(Y_0, Y_0) - C_0' \Sigma^{-1} C_0. \tag{2.31}$$

Matheron (1971) designou este método de predição linear por **kriging simples** visto pressupor o conhecimento do valor médio [16] (e da covariância).

Note-se ainda que, se impusermos a condição suplementar, $\sum_i l_i = 1$, na minimização de (2.29), então k_0 será nulo.

Suponhamos que agora se pretende calcular o melhor preditor linear do tipo (2.28) sujeito à condição

$$\sum_i l_i = 1,$$

[16] Alguns autores reservam o nome de kriging simples para o caso em que μ, embora desconhecido, é identificado com o seu estimador $\widehat{\mu} = \sum y_i / n$.

e que, além disso, o processo tem valor médio constante μ, desconhecido.

Já sabemos que nesse caso $k_0 = 0$, e portanto, a equação a minimizar é

$$E\Big[\big(Y(\boldsymbol{s}_0) - \sum_i l_i Y(\boldsymbol{s}_i)\big)^2\Big] = E\big[\big(\sum_{i=0}^{n} a_i Y(\boldsymbol{s}_i)\big)^2\big]$$

sujeita à condição $\sum a_i = 0$, ($a_0 = 1$ e $a_i = -l_i$).

Se o processo for intrinsecamente estacionário este valor médio é, de acordo com a Proposição (2.1), igual a

$$-\sum_i \sum_j a_i a_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j), \tag{2.32}$$

o que explica a razão pela qual, historicamente, o variograma apareceu em geoestatística no contexto do kriging. De facto, os coeficientes l_i do estimador óptimo, chamados **pesos de kriging**, encontram-se minimizando condicionalmente a expressão (2.32) sendo, portanto, funções do variograma $\gamma(\boldsymbol{h})$.

Evidentemente que se o processo for estacionário de segunda ordem, o preditor pode escrever-se em função do correlograma.

Este outro tipo de predição linear que acabámos de descrever designa-se por **kriging ordinário** e, tal como o kriging simples, pressupõe o conhecimento dos valores do covariograma (ou do variograma) para os pontos observados. Evidentemente que, na prática, estes são desconhecidos tornando-se necessário usar uma estimativa do covariograma quando se pretende fazer predição.

Esta situação dá-se não só com os preditores lineares como também com os não lineares: usualmente é necessário estimar alguns parâmetros dos quais depende a distribuição conjunta de $(Y_0, \boldsymbol{Y}')$ para se poder construir o preditor. O procedimento normal consiste em usar estimativas de máxima verosimilhança em substituição dos parâme-

tros desconhecidos, obtendo-se, não os preditores óptimos, mas sim estimativas de máxima verosimilhança desses preditores.

Na maioria dos casos não é possível conhecer a distribuição conjunta de $(Y_0, \boldsymbol{Y}')$ a não ser que se façam hipóteses simplificadoras sobre o processo subjacente. A mais comum é a de tomar Y como um processo gaussiano, à qual se juntam, frequentemente, as hipóteses de estacionaridade de segunda ordem e de isotropia. Este modelo é muito vantajoso pois a distribuição conjunta pretendida é ainda gaussiana e o preditor óptimo é linear em $\boldsymbol{Y}$.

Por todas estas razões vamos estudar mais em pormenor este modelo distinguindo o caso em que o valor médio é constante do caso em que isso não acontece.

Caso 1: Modelo gaussiano com valor médio constante μ.

Neste caso, o modelo para n localizações pode ser escrito na forma matricial,

$$\boldsymbol{Y} = \mu + \boldsymbol{\delta},$$

em que $\boldsymbol{\delta} = (\delta(\boldsymbol{s}_1), \ldots, \delta(\boldsymbol{s}_n)) = (\delta_1, \ldots, \delta_n)$ tem a distribuição

$$\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \Sigma).$$

O vector $\mathbf{0}$ tem todos os elementos nulos e, tal como já fizemos anteriormente,

$$\Sigma = \tau^2 I + \sigma^2 H(\phi) = \sigma^2 V,$$

com H matriz de elementos $\rho(\phi, \|\boldsymbol{s}_i - \boldsymbol{s}_i\|) = \rho(\phi, d_{ij})$.

Considere-se o vector coluna das observações do processo aumentado da observação em $\boldsymbol{s}_0$ desconhecida, $(Y_0, Y_1, \ldots, Y_n) = (Y_0, \boldsymbol{Y}')$, subdividido em dois vectores $\boldsymbol{Y}_1 = Y_0$ e $\boldsymbol{Y}_2 = \boldsymbol{Y}$.

De acordo com resultados já demonstrados, o estimador óptimo no sentido do erro quadrático médio será o valor médio condicional $E[\boldsymbol{Y}_1 | \boldsymbol{Y}_2]$.

É bem conhecido que a distribuição de $\boldsymbol{Y}_1|\boldsymbol{Y}_2$ continua a ser gaussiana com

$$E[\boldsymbol{Y}_1|\boldsymbol{Y}_2] = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{Y}_2 - \boldsymbol{\mu}_2),$$

e variância dada por

$$\Sigma(\boldsymbol{Y}_1|\boldsymbol{Y}_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21},$$

em que $\boldsymbol{\mu}_i = E[\boldsymbol{Y}_i]$, $i = 1, 2$, e a matriz Σ está subdividida em

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Adaptando a notação à situação em apreço temos,

$$\boldsymbol{\mu}_1 = \mu, \quad \boldsymbol{\mu}_2 = \mu\mathbf{1}, \quad \Sigma_{11} = \sigma^2 + \tau^2, \quad \Sigma_{22} = \sigma^2 V,$$

e

$$\Sigma_{21}' = \Sigma_{12} = \sigma^2\big(\rho(d_{01}), \ldots, \rho(d_{0n})\big) = \sigma^2\boldsymbol{r}',$$

o que conduz ao preditor

$$E[Y_0|\boldsymbol{Y}] = \mu + \boldsymbol{r}'V^{-1}(\boldsymbol{Y} - \mu\mathbf{1}), \tag{2.33}$$

com variância de predição

$$\mathrm{Var}(Y_0|\boldsymbol{Y}) = \tau^2 + \sigma^2(1 - \boldsymbol{r}'V^{-1}\boldsymbol{r}),$$

que devem ser comparados respectivamente com (2.30) e (2.31).

Já chamámos a atenção que, no caso gaussiano, o preditor de Y_0 pode escrever-se de forma linear nas observações e, de facto, é fácil verificar que a expressão (2.33) pode escrever-se como

$$\begin{aligned} E[Y_0|\boldsymbol{Y}] &= \mu + \sum_{i=1}^{n} l_i(\boldsymbol{s}_0)(Y_i - \mu) \\ &= \mu\big(1 - \sum_{i=1}^{n} l_i(\boldsymbol{s}_0)\big) + \sum_{i=1}^{n} l_i(\boldsymbol{s}_0)Y_i. \end{aligned}$$

Note-se que esta expressão mostra que o preditor de Y_0 é verdadeiramente um compromisso entre o valor médio do processo e as observações, compromisso esse que depende da posição de $\boldsymbol{s}_0$ e das localizações observadas, e também do modelo da covariância e dos seus parâmetros, através do que chamámos pesos da predição.

Evidentemente que, no cenário mais realista em que os parâmetros com que estamos a trabalhar são desconhecidos, devemos usar a estimativa do preditor óptimo de que já falámos anteriormente, ou seja o estimador que se obtém substituindo nas expressões anteriores os parâmetros desconhecidos por estimativas.

No caso do kriging ordinário μ é substituído pela sua estimativa dos mínimos quadrados generalizados,

$$\widehat{\mu} = (\mathbf{1}'V^{-1}\mathbf{1})^{-1}\mathbf{1}'V^{-1}\boldsymbol{Y},$$

onde ainda se está a supor que os parâmetros da covariância são conhecidos.

Caso 2: Modelo gaussiano com variáveis explicativas

Neste caso supõe-se que se podem conhecer os valores de p variáveis explicativas em qualquer dos pontos do domínio $\mathcal{D}$.

O modelo toma agora a forma mais geral

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\delta}$$

em que $\boldsymbol{\delta}$ é definido como no modelo anterior, X é a matriz das observações de p covariáveis em n localizações amostradas e $\boldsymbol{\beta}$ é o vector dos parâmetros da regressão da tendência espacial[17].

Com este modelo, o preditor passa a ter a expressão

$$E[Y(s_0)|\boldsymbol{Y}] = \boldsymbol{X}_0\boldsymbol{\beta} + \boldsymbol{r}'V^{-1}(\boldsymbol{Y} - \boldsymbol{X}\beta),$$

[17]Note-se que o modelo anterior é o caso particular deste com $p = 1$, X vector coluna unitário n-dimensional e $\boldsymbol{\beta}$ escalar μ.

mas a variância fica inalterada.

Este tipo de predição toma o nome de **kriging universal**.

Evidentemente que nada impede que façamos a predição de um valor que foi de facto observado, usando como preditores os valores observados nas outras localizações, e uma questão que se coloca naturalmente é a de saber se o valor estimado virá igual ao valor observado.

A resposta a esta questão é positiva se $\tau^2 = 0$, ou seja, se não existe efeito de pepita, e será negativa em caso contrário, o que mostra o importante papel que o efeito de pepita tem na predição e a consequente necessidade deste valor ser estimado com precisão.

Examinemos a questão da utilização prática deste preditor.

Uma vez que, neste caso, os parâmetros da tendência espacial são desconhecidos, como também acontece com os parâmetros do covariograma, teremos que os substituir pelos estimadores de máxima verosimilhança ou, caso estes não estejam disponíveis, por estimadores calculados por um outro qualquer método.

Faz-se notar que Cressie (1993) defende que, desde que $\boldsymbol{\beta}$ e $\boldsymbol{\theta}$ sejam estimados de forma eficiente, o preditor de Y_0 não deve ser muito afectado pela substituição dos verdadeiros valores pelas estimativas. Isso acontece mesmo no caso em que existe uma má especificação do modelo em termos da divisão entre efeitos de pequena e grande escala pois os dois estimadores tendem a compensar-se um ao outro não influenciando muito o resultado final da predição.

Consideremos agora o problema de predição de uma realização do processo em todo o domínio, ou seja, da construção de uma superfície predita.

Já chamámos a atenção para que o valores preditos dependem da função de correlação que se está a utilizar para modelar a associação

espacial entre observações. Se tivermos também em conta os resultados já provados sobre a ligação entre o correlograma e o alisamento das superfícies de realização do processo, é fácil concluir que a escolha de $\rho(.)$ vai influenciar o alisamento da superfície predita. De facto, uma escolha conveniente pode assegurar que estas sejam contínuas ou várias vezes diferenciáveis em média quadrática.

As Figuras 2.12 e 2.13 que se seguem, procuram ilustrar como é forte a dependência de que estamos a falar.

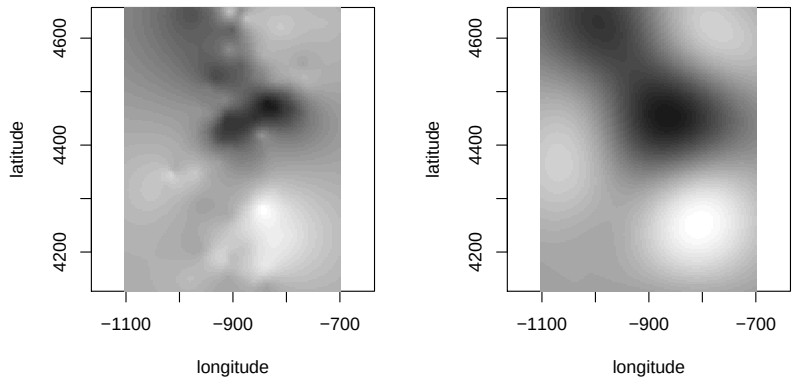


Figura 2.12 *Superfícies preditas para os modelos exponencial (à esquerda) e gaussiano (à direita) utilizando estimativas de mínimos quadrados ponderados para os parâmetros do covariograma.*

Estas figuras dizem de novo respeito ao exemplo da precipitação de Novembro de 2006 em Portugal continental.

Em qualquer delas apresentam-se superfícies preditas estimadas da raiz quadrada da precipitação cobrindo todo o território nacional. Nas coordenadas que estamos a utilizar, a longitude e a latitude dos pontos observados, cuja localização se pode consultar na Figura 1.3, variam respectivamente dentro dos intervalos $[-1049.35, -728.54]$ e $[4128.890, 4655.40]$.

O método de predição utilizado é sempre o kriging ordinário, mas na Figura 2.12 usam-se, para substituir os parâmetros desconhecidos do variograma, as estimativas obtidas pelo método de mínimos quadrados ponderados apresentadas na Tabela 2.1, enquanto que na Figura 2.13 se usam as estimativas de máxima verosimilhança restrita da Tabela 2.2.

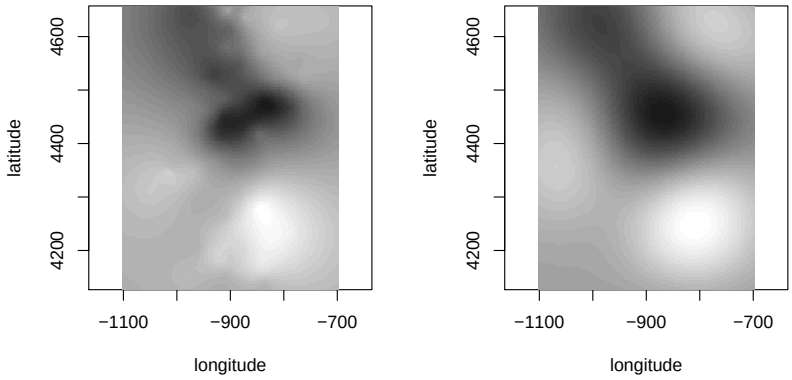


Figura 2.13 *Superfícies preditas para os modelos exponencial (à esquerda) e gaussiano (à direita) utilizando estimativas de máxima verosimilhança restrita para os parâmetros do covariograma.*

Em qualquer das figuras é patente a diferença de alisamento das superfícies preditas sendo este muito mais acentuado quando se utiliza o modelo gaussiano para o correlograma.

2.4.1 Considerações finais

Neste capítulo procurou dar-se uma visão global das técnicas clássicas de análise, modelação, estimação e predição usadas na geoestatística. Procurou também manter-se uma certa "leveza" na exposição dos conceitos e demonstrações sem contudo abandonar completamente a sua fundamentamentação matemática. A exposição segue os passos

que devem ser normalmente dados quando se pretende fazer predição a partir da observação de um fenómeno num certo conjunto de localizações e, por essa razão, o texto é acompanhado por um exemplo real com o qual se ilustra cada uma das fases desse processo.

O estudo do exemplo foi levado a cabo através do pacote `geoR` do *software* `R`, `http://www.leg.ufpr.br/geoR`, cuja utilização recomendamos pela simplicidade de manuseamento e pela versatilidade e profundidade das análises que com ele se podem realizar. Este pacote foi desenvolvido por Peter Diggle e Paulo Ribeiro Jr., cujo livro "Model-based Geostatistics" (2007) também recomendamos aos leitores interessados em aprofundar esta matéria.

A exposição não tem a pretensão de ser exaustiva e muitos tópicos são apenas aflorados de forma a que o leitor interessado possa depois seguir apresentações mais completas em textos para eles especialmente dirigidos ou em livros como o de Noel Cressie, " Statistics for Spatial Data" (1993) onde todos os temas que abordámos são desenvolvidos com mais robustez e profundidade.

Por último não podemos deixar de referir que, mais recentemente, a análise bayesiana de dados espaciais tem proporcionado uma nova visão dos problemas clássicos da geoestatística que alguns livros, como o de Diggle e Ribeiro Jr. (2007) que já referimos e o de Sudipto Banerjee, Bradley Carlin e Alan Gelfand (2004), " Hierarchical Modeling and Analysis for Spatial data", expõem com grande clareza.

Capítulo 3

Modelos Referentes a Áreas

3.1 Questões básicas sobre dados referentes a áreas

Como dissemos na introdução, os dados que vamos estudar neste capítulo dizem respeito a observações em áreas que subdividem uma certa região limitada $\mathcal{D}$. Essas áreas constituem uma partição de $\mathcal{D}$ e podem estar relacionadas com regiões geográficas, como por exemplo regiões administrativas, sendo na maioria desses casos irregulares, ou serem células de área idêntica dispostas numa grelha regular, como por exemplo os píxeis de uma imagem.

Normalmente usaremos a designação de **blocos** para nos referirmos genericamente a qualquer dos tipos de áreas de que falámos, mas consoante as aplicações também é vulgar usar o nome de **células** ou **regiões** ou **áreas**, como aliás acabámos de fazer.

Supondo que a partição é constituída pelos blocos $B_i,\ i = 1, \ldots, n$, designaremos por $\boldsymbol{Y} = \big(Y(B_1), \ldots, Y(B_n)\big)' \in \mathbb{R}^n$ as observações efectuadas nos correspondentes blocos, que tipicamente dizem respeito a somas ou médias de uma certa variável aí.

Frequentemente cada bloco tem um ponto representante que assume um papel importante em alguns dos modelos que vamos utilizar

para medir a proximidade dos blocos. Esses pontos são designados por **centróides** visto que, na maior parte dos casos, as coordenadas da sua posição são as dos pontos médios dos intervalos entre o mínimo e o máximo para cada uma das coordenadas de todos os pontos do bloco. Embora não seja muito vulgar, $\boldsymbol{Y}$ pode medir, não um valor acumulado ou médio da variável de interesse nos blocos, mas sim o valor dessa variável exactamente nos pontos representantes que se escolheram de acordo com uma regra especial.

Exemplos de dados referentes a áreas são o número de casos de uma determinada doença ou o rendimento médio dos agregados familiares por região administrativa ou postal de um certo país, a proporção de área urbana em cada quadrícula de 10km de lado de um reticulado que cobre uma certa região geográfica ou ainda o tipo de ocupação do solo (variando dentro de uma classificação pré-estabelecida) do centro de cada uma dessas quadrículas.

Com exemplos tão variados é evidente que serão muito diversificados os objectivos que se pretendem atingir com a análise de dados referentes a áreas, no entanto, a maioria dos problemas pode classificar-se dentro de alguns tipos gerais que vamos passar a apresentar.

- Em muitos casos os valores da variável em apreço não parecem estar espalhados ao acaso. Existem áreas do domínio onde os blocos apresentam valores semelhantes mais elevados que a média e outras onde parece acontecer o contrário. A esta distribuição desigual chama-se **padrão espacial** e numa primeira fase a detecção da sua existência é feita através da observação de mapas.

Os mapas usados nesta análise são mapas onde figuram os blocos com os correspondentes valores da variável inscritos, ou mapas coropletos, de que já apresentámos um exemplo na Figura 1.4, onde são pintados no mesmo tom os blocos que apresentam valores da variável dentro de cada um dos intervalos de um conjunto de intervalos

contíguos que cobrem todos os valores observados da variável de interesse.

Esta análise visual é muito falível por várias razões de que teremos oportunidade de falar, mas das quais destacamos já o facto de, mesmo com medidas independentes em cada bloco, se poder esperar algumas manchas de valores semelhantes. Como tal, a quantificação do que intuitivamente divisamos como um padrão espacial, e que se percebe estar ligado à existência de uma associação espacial entre blocos que estão perto no espaço, terá que ser feita de forma a que se possa detectar sem ambiguidade a sua existência. O nosso objectivo neste campo vai ser o de apresentar medidas desta associação e apresentar testes que permitam verificar se os seus valores são significativos.

- Frequentemente também aparece um grande número de blocos com valores muito elevados e outros com valores muito baixos.

Embora a existência de valores extremos seja esperada mesmo em amostras de observações independentes, nos dados referentes a áreas estes parecem ser em número e tomar valores excepcionalmente maiores ou menores do que o que seria de esperar por obra do acaso. Muitas vezes, isso resulta das fronteiras entre blocos serem linhas de divisão artificiais do domínio que podem fazer aglomerar, também artificialmente, certos fenómenos em alguns dos blocos, rareando-os noutros. Veja-se o exemplo de uma freguesia que tendo um lar de idosos recebe das freguesias vizinhas os elementos da população que são mais atreitos a certas doenças, nomeadamente as que têm a idade como factor de risco. O número de doentes cardíacos pode ser muito elevado nessa freguesia e extremamente baixo nas freguesias vizinhas exclusivamente por esta razão. Por outro lado, contagens de acontecimentos raros referentes a períodos pequenos têm, como se sabe, propensão a apresentar muitos zeros.

Desta forma, os dados observados podem não dar uma ideia cor-

recta da distribuição do fenómeno no domínio, e um alisamento destes "bicos" discrepantes aproximando-os da média proporciona uma visão do fenómeno mais compatível com a realidade.

Este alisamento pode tomar muitas formas. Por vezes, caso estejamos a contar ocorrências de um fenómeno na população residente no domínio, basta padronizar os dados por classe etária e género para conseguir uma ideia mais correcta da distribuição do fenómeno, e algum alisamento dos dados, mas na maioria das vezes este tipo de métodos não resolve o problema e é necessário aplicar outras técnicas.

Evidentemente, que entre a apresentação dos dados tal qual e o alisamento máximo que seria conseguido atribuindo a média das observações a todos os blocos, algum equilíbrio tem de ser encontrado. A extensão e a forma de alisar pode ser definida deterministicamente, por métodos algorítmicos como os que estão disponíveis na maior parte dos pacotes de *software* SIG, ou usando modelos estatísticos que nos interessam especialmente e sobre os quais falaremos também neste capítulo embora de forma breve. Muitos destes modelos são mais interessantes se forem examinados no contexto da regressão, ou seja, usando variáveis que podem explicar grande parte da variação das respostas entre blocos e incorporando nos resíduos uma estrutura espacial que explica a variação restante. Evidentemente que o objectivo será o de explicar através de covariáveis a maior parte da variação entre blocos, dado que a estrutura espacial não é realmente uma explicação física do fenómeno, mas sim o resultado da presença de covariáveis que ainda não foram identificadas.

- Embora o âmbito deste curso não permita abordar outros problemas além dos que acabámos de apresentar, não podemos deixar de referir um dos que mais aplicação tem em situações da vida real. Suponhamos que dos blocos observados alguns são subdivididos e outros agregados. Como poderão ser encontradas estimativas para es-

sas novas áreas? Este problema, conhecido por *modifiable areal unit problem* (MAUP), pode também ter uma abordagem determinística, baseada essencialmente na proporcionalidade do fenómeno em causa à área dos blocos, ou estocástica sobre a qual recomendamos a leitura de Mugglin e Carlin (1998), Mugglin *et al.*(2000) e ainda livro de Banerjee *et al.*(2004).

3.2 Padrão espacial

Uma análise exploratória dos dados é sempre um passo necessário para se iniciar o estudo da presença de algum padrão espacial. Essa análise exploratória inclui a usual observação da distribuição dos dados independentemente da sua localização e também o seu mapeamento.

Já referimos que a forma mais usual de mapear os dados é através de mapas coropletos, instrumento cuja construção tem problemas análogos aos da construção de histogramas. De facto, o número de intervalos a escolher não deve ser demasiadamente grande, sob pena de termos alguns intervalos pouco representados e um gradiente de cores ou tons demasiadamente grande para ser apercebida pelo olho humano. Um número entre 5 e 8 (nunca devendo exceder 10) considera-se normalmente adequado.

Não queremos deixar de chamar a atenção para a facilidade com que estes mapas são manipuláveis, mesmo considerando que as suas fronteiras foram naturalmente e não artificialmente criadas. Através do uso do gradiente de cores e dos limites dos intervalos pode ser dada a ideia de presença muito forte de valores elevados ou o contrário: no primeiro caso basta usar cores ou tons escuros (a progressão no escurecimento de uma cor está psicologicamente ligada ao agravamento de uma situação) para os valores elevados e dividir os intervalos de

forma a que grande parte dos blocos seja pintada de escuro.

Por estas razões, o uso dos coropletos deve servir apenas para identificar possíveis aglomerações de blocos cuja verdadeira existência deve depois ser confirmada por outros métodos mais formais que passamos a estudar. No entanto, para o fazermos necessitamos de introduzir uma ferramenta que também vai ser útil na modelação destes dados.

Trata-se da **matriz de vizinhanças** ou **matriz de pesos** notada habitualmente pela letra W inicial de *weights*, cujas entradas vão medir a ligação espacial entre os blocos.

Usualmente, $w_{ii} = 0$, $i = 1, \ldots, n$, distância mínima entre os blocos, mas para $i \neq j$, w_{ij}, medida da associação espacial entre o bloco i e o bloco j, pode ser definida de muitas maneiras diferentes.

Muitas destas formas são de escolha binária e, nesse caso, é mais usual W ter a designação de matriz de vizinhanças porque se consideram como **vizinhos** os blocos i e j para os quais $w_{ij} = 1$. Algumas dessas escolhas são mais comuns:

1. **Vizinhança por adjacência**

   São iguais a 1 todos os elementos w_{ij} para os quais os blocos correspondentes partilham uma fronteira. Todos as outras entradas da matriz são nulas.

   Por vezes, se os blocos são em quadrícula consideram-se também como vizinhos os blocos que partilham um vértice.

2. **Vizinhanças baseadas na distância entre centróides**

   Uma definição bastante usual é a de considerar como vizinhos todos os blocos cuja a distância entre os respectivos centróides seja inferior a uma distância fixa d.

   Uma outra definição possível é a de considerar como vizinhos de um bloco, os m blocos mais perto de si, tomando de novo a distância entre blocos como a distância entre os respectivos

centróides.

Outras formas para as entradas da matriz de pesos dizem respeito a escolhas não binárias.

Uma das mais comuns é notada por $\widetilde{W}$ e é construída á custa de qualquer matriz de pesos estandardizando-a por linhas. As entradas desta matriz são dadas por

$$\widetilde{w}_{ij} = w_{ij}/w_i^+ \quad \text{em que} \quad w_i^+ = \sum_{j=1}^{n} w_{ij},$$

o que faz com que $\widetilde{W}$ seja estocástica por linhas

$$\widetilde{W}\mathbf{1} = 1.$$

No caso da matriz ser de vizinhanças w_i^+ corresponde ao número de vizinhos de i.

Outras escolhas baseiam os pesos em funções não binárias do inverso da distância entre centróides, uma vez que há situações em que é natural atribuir um valor mais elevado à associação entre blocos que se encontram mais próximo.

Convém notar que os pesos podem ser atribuídos com base noutro tipo de características que não a distância entre centróides, como por exemplo, a proporção de fronteira partilhada pelos blocos, ou um índice de acessibilidade construído a partir da quantidade de vias de comunicação entre blocos.

Note-se que alguns dos tipos de matrizes que apresentámos podem conduzir, para blocos com disposição irregular, a matrizes não simétricas. Apesar desta situação ser admissível, na maioria dos casos consideram-se relações de vizinhança simétricas.

Faz-se notar ainda que as matrizes de vizinhança são muitas vezes encaradas como matrizes de adjacência de um grafo dirigido cujos

vértices são os centróides, que estarão unidos por arestas dirigidas caso os blocos sejam vizinhos. A estrutura de vizinhanças que esta matriz define tem a função de introduzir formalmente nos modelos que vamos usar a correspondente estrutura da associação espacial entre blocos.

Em vez de construir uma só matriz de vizinhanças para modelar a estrutura de associação espacial pode construir-se um sistema de matrizes que consideram diversos tipos de vizinhos. No que tem mais interesse as matrizes dizem respeito a vizinhos que vão estando progressivamente mais longe. A sua construção assenta na divisão do intervalo que contém todas as distâncias entre centróides em subintervalos contíguos

$$(0, d_1], (d_1, d_2], \ldots, (d_{k-1}, d_k].$$

Chamamos vizinhos de ordem i aos blocos que tenham centróides a uma distância incluída no intervalo $(d_{i-1}, d_i]$. Na correspondente matriz de vizinhanças de ordem i, $W^{(i)}$ todas as entradas que dizem respeito a vizinhos dessa ordem são iguais a 1 sendo nulas todas as outras.

Estas matrizes de pesos vão ter um papel essencial na definição das medidas de associação espacial que vamos tratar na próxima secção.

3.2.1 Medidas de associação espacial

As duas estatísticas mais conhecidas para medir a associação espacial de dados referentes a pontos são: estatística I de Moran, Moran (1948), e a estatística c de Geary, Geary(1954). Elas são dadas respectivamente pelas seguintes funções dos dados e da estrutura de vizinhanças:

$$I = \frac{n \sum_i \sum_j w_{ij}(Y_i - \overline{Y})(Y_j - \overline{Y})}{\left(\sum_{i \neq j} w_{ij}\right) \sum_i (Y_i - \overline{Y})^2}$$

$$c = \frac{(n-1)\sum_i \sum_j w_{ij}(Y_i - Y_j)^2}{2\big(\sum_{i \neq j} w_{ij}\big)\sum_i (Y_i - \overline{Y})^2},$$

em que w_{ij} são as entradas de uma matriz de pesos[1]. Antes de passarmos ao estudo mais pormenorizado destas estatísticas chama-se a atenção que o I de Moran e o c de Geary podem ser encarados, respectivamente, como análogos para dados referentes a áreas das medidas de associação espacial para dados referentes a pontos, usadas no correlograma e semivariograma amostrais.

Por essa razão usa-se por vezes um gráfico construído no mesmo espírito do correlograma para apreciar a forma como a associação espacial entre blocos varia com a "distância". Trata-se de um gráfico onde se dispõe os valores de I_i calculados com as matrizes de vizinhança de ordem i contra os valores de i, $i = 1, \ldots, k$. Se existir padrão espacial, espera-se, por analogia com o correlograma, que o gráfico decresça inicialmente de forma rápida e que estabilize perto de 0.

Teste global de Moran

Para poder utilizar a estatística I em testes teríamos que saber a sua distribuição sob a hipótese nula das observações serem independentes. No entanto, a não ser para conjuntos de observações muito pequenos, o cálculo exacto desta distribuição é impossível sendo por isso necessário encontrar uma sua aproximação que, como vamos ver, pode ser obtida sob diferentes hipóteses. Para começar, formulemos a hipótese nula de dois modos diferentes,

Hipótese $\mathbf{H_0}$: As observações são tiragens independentes de uma distribuição normal,

[1] Inicialmente estas estatísticas foram definidas para matrizes de vizinhanças mas Cliff e Ord (1973) generalizaram-nas para o caso de matrizes de pesos.

Hipótese $\mathbf{H}_0'$: As observações são tiragens independentes de uma distribuição desconhecida,

às quais juntamos ainda uma terceira hipótese que se vai revelar muito útil,

Hipótese $\mathbf{H}_0''$: Um conjunto de dados observados é uma tiragem feita com igual probabilidade do conjunto das $n!$ permutações possíveis dos dados.

Para condições não muito restritivas, a distribuição de I, quando n aumenta, tende para a distribuição gaussiana sob qualquer das hipóteses nulas.

O resultado com mais consequências (Cliff e Ord, 1981) diz, no entanto, respeito aos dois primeiros momentos das distribuições da estatística sob as três hipóteses. De facto, ficando a distribuição gaussiana completamente definida por estes momentos, basta saber quais são as correspondentes expressões nos três casos e aplicar limites para se conhecer, para qualquer deles, a distribuição limite de I, como é comprovado pelo seguinte teorema:

Teorema 3.1. *Seja $F(y)$ uma distribuição determinada de forma única pelos seus momentos, μ_k, e seja $\{F_n(y)\}_n$, $n = 1, \ldots$ uma sucessão de funções distribuição com momentos μ_{k_n}. Se $\mu_{k_n} \to \mu_k$ quando $n \to \infty$, para $k = 1, 2, \ldots$, então $F_n(y)$ converge para $F(y)$ em todos os pontos de continuidade de $F(y)$.*

O resultado de que falamos é o seguinte:

Teorema 3.2. *Para qualquer distribuição subjacente aos dados, desde que tenha variância finita, é válido o seguinte resultado*

$$E(I) = -1/(n-1) \quad e \quad E(I^2) = E\big[E_R[I^2]\big],$$

em que $E_R[\cdot]$ designa o valor esperado sob H_0''.

Isto quer dizer que o valor esperado de I^2, sob a hipótese da permutação aleatória, fornece uma estimativa não enviesada de $E[I^2]$ e, portanto, da variância de I, qualquer que seja a distribuição subjacente. Desta forma, quando queremos aproximar a distribuição de I, podemos usar $E_R[I^2]$ em vez de $E[I^2]$ se desconhecermos a distribuição subjacente aos dados.

Uma segunda consequência deste resultado é também importante.

Note-se que, embora a distribuição assintótica de I seja normal, para valores relativamente pequenos de n a distribuição de I pode estar ainda bastante distante desta. De facto, a forma da distribuição depende não só de n mas também do formato dos blocos e do número médio de vizinhos por bloco, dos pesos w_{ij} e da distribuição de Y. No entanto, se pudessemos construir a distribuição de I a partir das $n!$ correspondentes permutações dos dados, ou seja, sob a hipótese H_0'', teríamos de acordo com o resultado, uma "boa"estimativa da distribuição de I sob as duas outras hipóteses.

Claro que, a construção desta distribuição é impraticável mesmo para valores não muito elevados de n, mas a utilização de métodos Monte Carlo permite construir uma distribuição aproximada a partir de um número escolhido de permutações (normalmente 999), que pode ser utilizada num teste da hipótese genérica de não existência de associação espacial.

Sob a hipótese H_0, o valor da variância é dado por

$$\mathrm{Var}(I) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)^2 S_0^2},$$

com $S_0 = \sum_{i \neq j} w_{ij}$, $S_1 = \frac{1}{2}\sum_{i \neq j}(w_{ij} + w_{ji})^2$ e $S^2 = \sum_k(\sum_j w_{kj} + \sum_i w_{ik})^2$.

Sob a hipótese H_0'' a expressão da variância é mais complexa (Cliff e Ord, 1981), mas em qualquer dos casos é interessante notar que, uma vez estabelecida a estrutura de pesos, ou seja, a matriz W, tanto

a variância como o valor médio da distribuição de I são invariantes face aos valores observados e também face à área e forma dos blocos, embora o mesmo não se possa dizer da forma da distribuição.

Também se deve notar que embora seja tentador interpretar I como um coeficiente de correlação, visto que de certa forma mede a correlação espacial dos dados (lembre-se a analogia que já referimos com os dados referentes a pontos), a verdade é que as usuais propriedades da correlação não se verificam para esta estatística. Por exemplo, a distribuição de I não está estritamente suportada pelo intervalo $[-1, 1]$. Em geral o limite superior de $|I|$ é inferior a 1, mas pode ser maior em casos especiais de distribuições muito irregulares das entradas da matriz dos pesos.

Evidentemente que dada a distância a que a distribuição verdadeira pode estar da assintótica, os resultados dos testes de existência de associação espacial ou os intervalos de confiança baseados nesta estatística devem ser encarados com as devidas precauções, no entanto, um valor da estatística de Moran significativamente positivo indica usualmente a presença de aglomerados constituídos por valores altos e outros por valores baixos, enquanto que um valor da estatística de Moran significativamente negativo indica que os valores altos e baixos têm um padrão axadrezado.

Outro teste para a associação espacial baseia-se na distribuição de I calculada via Monte Carlo. Trata-se de calcular o valor p empírico da hipótese H_0, verificando a posição da estatística de teste na amostra ordenada dos valores de I calculados a partir de cada uma das réplicas, e compará-lo em seguida com um valor crítico como por exemplo $\alpha = 0.05$.

Para os dados da mortalidade por cancro de estômago apresentados na Figura 1.4 calculámos, para todos os concelhos, os valores da razão de mortalidade padronizada (SMR) por género e classe etária, em que

a mortalidade padrão foi a mortalidade por cancro de estômago da população do continente.

A Figura 3.1 apresenta o mapa coropleto da razão de mortalidade padronizada, sendo interessante comparar este com o mapa para o número de casos que acabámos de referir e verificar como é diferente a distribuição dos valores elevados e baixos nos dois casos. Isso é uma consequência directa da grande variação da distribuição da população portuguesa por sexo e classe etária entre os concelhos do continente.

A figura parece mostrar a presença de um padrão espacial com valores mais elevados de risco relativo aglomerados no norte e no sul de Portugal e valores mais baixos no centro. A confirmação da existência de associação espacial fez-se através do teste global de Moran cujos resultados também apresentamos mais abaixo.

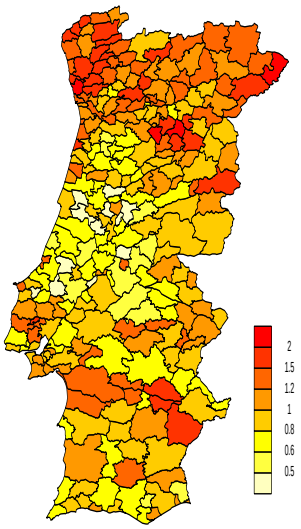


Figura 3.1 *Mapa da razão de mortalidade padronizada (SMR) para o cancro de estômago nos concelhos de Portugal continental entre 1991 e 1995.*

A Figura 3.2 apresenta o histograma da distribuição empírica de I, calculada via Monte Carlo a partir de 999 réplicas dos dados originais. A matriz de vizinhanças para os concelhos de Portugal foi construída à custa da vizinhança por adjacência.

A linha tracejada vertical representa a posição da estatística de teste que tem o valor $I = 0.452$. Como esse valor é superior ao alcançado para todas as réplicas, o valor p correspondente é 0.001 o que apoia fortemente a hipótese da presença de um padrão espacial. O teste efectuado sob a hipótese H_0' (a normalidade não se verifica para estes riscos relativos) aponta na mesma direcção, visto que o valor p vem também extremamente baixo ($< 10^{-15}$).

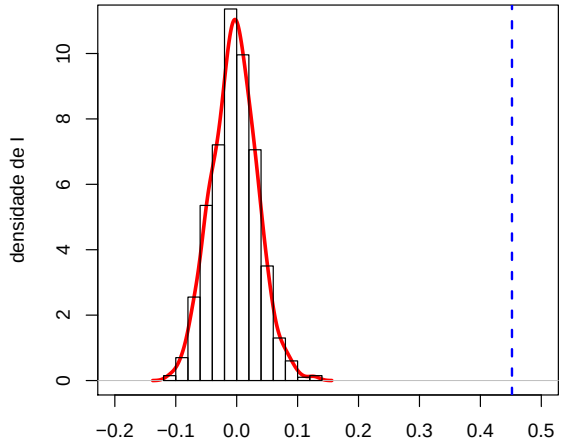


Figura 3.2 *Histograma da distribuição empírica de I calculada com 999 réplicas e linha vertical no valor da estatística de Moran referente às observações da razão de mortalidade padronizada para o cancro de estômago .*

Teste de Geary

Para a estatística c de Geary já apresentada e para o correspondente teste, poderíamos fazer deduções análogas às que acabámos de fazer e que, por isso, não repetimos. O leitor interessado poderá segui-las em Cliff e Ord (1981). Refere-se apenas que estatística é sempre positiva, de valor médio unitário e que valores inferiores ao valor médio indicam associação espacial positiva. Sob a hipótese H_0 a variância do estimador é

$$\mathrm{Var}(c) = \frac{(2S_1 + S_2)(n-1) - 4S_0^2}{2(n+1)S_0^2},$$

e a correspondente expressão para variância sob a hipótese H_0'' pode ser consultada na referência anterior.

Segundo indicam alguns autores (Cliff e Ord (1969)) a variância de I é menos afectada pela distribuição subjacente aos dados do que a variância de c o que faz com que a utilização de c seja menos frequente quando se pretende efectuar testes.

Gráfico de Moran

O gráfico de Moran é um gráfico criado por Luc Anselin (1988) para explorar visualmente a autocorrelação espacial.

Suponhamos que, para o conjunto de blocos em observação está definida uma matriz de pesos. O gráfico de Moran tem como abcissas os valores estandardizados das observações Y_i e em ordenadas os valores da média ponderada por w_{ij} dos valores estandardizados de todos os outros elementos. A unidade de medida dos eixos é, como é óbvio, o desvio padrão das observações.

A utilização deste gráfico é mais vulgar quando a matriz de pesos é uma matriz de vizinhanças, caso em que as ordenadas são apenas a média dos valores estandardizados dos vizinhos de cada bloco.

Note-se que os pontos dos quadrantes ímpares do gráfico correspondem a blocos positivamente correlacionados com eventuais aglomerados, se a valores muito elevados (baixos) das abcissas correspondem valores também muito elevados (baixos) das ordenadas. De facto, isso quer dizer que os blocos com observações elevadas tendem a estar rodeados de blocos com observações elevadas e que o análogo acontece para observações baixas.

Os pontos dos quadrantes pares correspondem a uma correlação negativa (eventuais outliers se estão longe da origem): no segundo quadrante estão pontos que tomando um valor baixo se encontram rodeados por vizinhos de valor elevado, acontecendo a situação inversa no quarto quadrante.

Desta forma se um gráfico de Moran tem os seus pontos essencialmente nos quadrantes ímpares ele indicia a presença de um padrão espacial com aglomerados de valores altos e baixos (correlação positiva). Se os valores estão essencialmente nos quadrantes pares, então deve existir um padrão espacial axadrezado (correlação negativa). Finalmente, se os pontos estão em nuvem em torno da origem, não deve existir padrão espacial.

Como exemplo de aplicação deste método visual de detecção de associação espacial apresentamos na Figura 3.3 o gráfico de Moran correspondente aos dados que temos vindo a tratar. A matriz W utilizada é, como anteriormente, a de vizinhanças por adjacência.

Uma vez que pelos métodos anteriores já foi detectada a presença de associação espacial positiva esperar-se-ia que os pontos estivessem essencialmente nos quadrantes ímpares, o que realmente se verifica.

Note-se que o declive da recta de regressão que também está representada no gráfico é o valor da estatística I de Moran.

Outras estatísticas do tipo Moran foram proposta por Luc Anselin

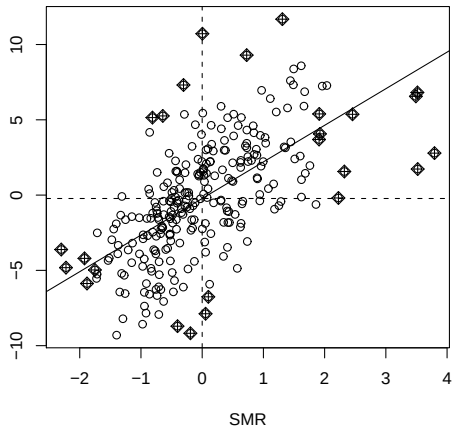


Figura 3.3 *Gráfico de Moran para a razão de mortalidade padronizada.*

(1995) para serem aplicadas a nível local. Elas são conhecidas sob o nome de LISA - *Local Indicators of Spatial Association* - e estão implementadas, juntamente com as que aqui apresentámos, no *software* gratuito `GeoDa` desenvolvido no Spatial Analysis Laboratory do Departamento de Geografia da Universidade de Illinois, Urbana-Champaign. No *software* `R` encontram-se outros pacotes adaptados ao estudo de dados referentes a áreas como é o caso do `spdep` que foi utilizado em todos os exemplos deste capítulo.

3.3 Alisamento espacial

Nesta secção vamos tratar do problema do alisamento das observações de cuja necessidade falámos no início deste capítulo.

Comecemos por notar que o mapa coropleto já corresponde, de certa forma, a uma resposta a este problema, visto que, ao dividir os dados num pequeno número de classes, os alisa. Já sabemos, contudo, os problemas associados a esta técnica.

Uma outra técnica determinística de alisamento, muito usada em *software* SIG onde toma o nome de "filtragem", consiste em usar como estimativa alisada para o bloco i uma combinação linear convexa do valor observado e da média pesada dos valores dos seus vizinhos,

$$\widetilde{Y}_i = (1-\alpha)Y_i + \alpha Y_i^*$$

em que $\alpha \in (0,1)$ e

$$Y_i^* = \sum_j \frac{w_{ij}Y_j}{w_i^+}.$$

Tal como no gráfico de Moran, se a matriz W é de vizinhanças, Y_i^* é apenas a média dos valores dos blocos vizinhos de i. O grau de alisamento é controlado pelo valor de α e quanto maior ele for maior será a aproximação à média conseguida. Note-se que casos extremos $\alpha = 0$ e $\alpha = 1$ foram retirados porque correspondiam, respectivamente, a não fazer qualquer alisamento ou a fazer um alisamento em que o próprio valor observado não participava na estimativa.

Passamos agora a estudar um outro método de alisamento, desta feita estocástico, especialmente dirigido ao alisamento de dados poissonianos.

Estes estimadores, propostos por Clayton and Kaldor (1987), são mais usados no âmbito da modelação de mortalidade/morbilidade de doenças raras e não contagiosas, com base no número observado de mortes/casos de doença em divisões da região de estudo, mas podem ser usados para qualquer outro tipo de incidências com as mesmas propriedades.

O seu método assume o seguinte modelo de Poisson para o número de casos observados no bloco i, $\{Y_i : \ i = 1, \ldots, n\}$, dados os riscos relativos $\{\theta_i, \ i = 1, \ldots, n\}$ e sendo $\{E_i : \ i = 1, \ldots, n\}$ os correspondentes números esperados de casos, supostos conhecidos:

$$Y_i \sim \text{Poisson}(E_i\theta_i), \qquad i = 1, \ldots, n \tag{3.1}$$

Dado que temos apenas uma observação por bloco, o estimador de máxima verosimilhança para o parâmetro θ_i baseado nessa observação seria, evidentemente, $\hat{\theta}_i = SMR_i = Y_i/E_i$ e o correspondente estimador do número de casos nesse bloco seria exactamente igual ao valor observado, o que não produz qualquer tipo de alisamento.

A ideia geral do método é a de procurar encontrar um estimador para θ_i que aproveite o conhecimento que temos do número de casos nos outros blocos o que é feito por estimação Bayes-empírica. Para isso assume-se que os riscos relativos têm uma distribuição *a priori* Gama(ν, α), de valor médio $\frac{\nu}{\alpha}$ e variância $\frac{\nu}{\alpha^2}$,

$$p(\theta_i) = \frac{\alpha^\nu}{\Gamma(\nu)} {\theta_i}^{\nu-1} \exp(-\alpha\theta_i), \quad \theta_i \geq 0;\ \alpha, \nu > 0; \quad i = 1, \ldots, n,$$

A distribuição *a posteriori* dos riscos relativos é então determinada como:

$$\begin{aligned} p(\theta_i|y_i) &\propto p(\theta_i)p(y_i|\theta_i) \propto {\theta_i}^{\nu-1} \exp(-\alpha\theta_i) \exp(-E_i\theta_i)(E_i\theta_i)^{y_i} \\ &\propto {\theta_i}^{y_i+\nu-1} \exp(-(E_i+\alpha)\theta_i), \end{aligned}$$

correspondendo a uma distribuição Gama$(y_i + \nu, E_i + \alpha)$.

O seu valor médio (esperança *a posteriori*) é o estimador proposto para o risco relativo θ_i:

$$\hat{\theta}_i = \frac{Y_i + \nu}{E_i + \alpha} \tag{3.2}$$

Claro que, para determinar completamente (3.2), os parâmetros *a priori* α e ν têm de ser conhecidos, o que não acontece usualmente na prática, levando assim a que tenham de ser também estimados.

Tal é conseguido maximizando a distribuição marginal dos dados Y_i, que se verifica ser Binomial Negativa com média $\frac{E_i\nu}{\alpha}$ e variância $\frac{E_i\nu}{\alpha} + \frac{E_i^2\nu}{\alpha^2}$, como se prova de seguida:

$$\begin{aligned}
p(y_i) &= \int_{\Omega_\theta} p(y_i|\theta_i)p(\theta_i)\,d\theta_i = \\
&= \int_0^{+\infty} \frac{\exp(-\theta_i E_i)(\theta_i E_i)^{y_i}}{y_i!}\frac{\alpha^\nu}{\Gamma(\nu)}\theta_i^{\nu-1}\exp(-\alpha\theta_i)\,d\theta_i = \\
&= \frac{E_i^{y_i}}{y_i!}\frac{\alpha^\nu}{\Gamma(\nu)}\int_0^{+\infty}\exp(-\theta_i(E_i+\alpha))\theta_i^{y_i+\nu-1}\,d\theta_i = \\
&= \frac{E_i^{y_i}}{y_i!}\frac{\alpha^\nu}{\Gamma(\nu)}\left(\frac{1}{E_i+\alpha}\right)^{y_i+\nu}\Gamma(y_i+\nu) = \\
&= \left(1-\frac{\alpha}{E_i+\alpha}\right)^{y_i}\left(\frac{\alpha}{E_i+\alpha}\right)^{\nu} C_{y_i}^{\nu+y_i-1}
\end{aligned}$$

Sendo, então, $Y_i \sim \text{Neg Bin}\left(\nu, \frac{\alpha}{E_i+\alpha}\right)$, do logaritmo da verosimilhança, por derivação, facilmente obtemos as equações de verosimilhança seguintes, a serem pois resolvidas pelos procedimentos iterativos padrão:

$$\frac{\hat{\nu}}{\hat{\alpha}} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i+\hat{\nu}}{E_i+\hat{\alpha}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i, \tag{3.3}$$

$$\sum_{i=1}^{n}\sum_{j=0}^{y_i-1}\frac{1}{\hat{\nu}+j} + n\log(\hat{\alpha}) - \sum_{i=1}^{n}\log(E_i+\hat{\alpha}) = 0. \tag{3.4}$$

Alternativamente, uma outra forma mais simples e computacionalmente mais fácil de fazer a estimação, é usar o estimador dos momentos para a variância dos parâmetros do risco $\{\theta_i\}$, $\frac{\nu}{\alpha^2}$.

Este é conseguido fazendo a seguinte estatística de qui-quadrado de Pearson, baseada nos momentos marginais das observações Y_i, igual ao seu valor médio assintótico:

$$\sum_{i=1}^{n} \left(\frac{y_i - \mathrm{E}[Y_i]}{\sqrt{\mathrm{Var}[Y_i]}} \right)^2 = n - 1,$$

de onde se deduz a igualdade

$$\frac{\hat{\nu}}{\hat{\alpha}^2} = \frac{1}{n-1} \sum_{i=1}^{n} \left(1 + \frac{\hat{\alpha}}{E_i} \right) \left(\hat{\theta}_i - \frac{\hat{\nu}}{\hat{\alpha}} \right)^2. \tag{3.5}$$

Assim as equações (3.3) e (3.5) podem ser usadas recursivamente para determinar $\hat{\alpha}$ e $\hat{\nu}$ num processo conhecido por *expectation-maximization algorithm*: O processo é iniciado escolhendo valores plausíveis para $\alpha^{(0)}$ e $\nu^{(0)}$ a partir dos quais se constroem, as correspondentes estimativas de Bayes empíricas do risco θ_i usando (3.2) (*expectation*). Seguidamente, usam-se as equações (3.3) e (3.5) para calcular estimativas actualizadas dos parâmetros α e ν (*maximization*) e o processo reinicia-se.

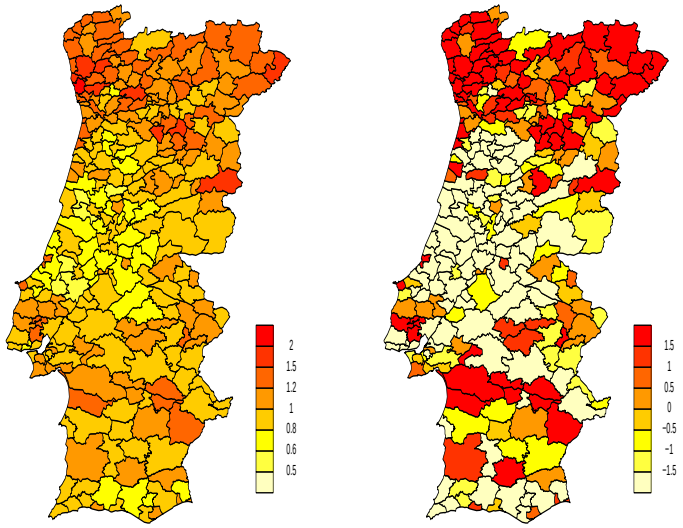
Note-se que os estimadores do risco assim obtidos

$$\hat{\theta}_i = \frac{Y_i + \widehat{\nu}}{E_i + \widehat{\alpha}}$$

são um compromisso entre os valores observados do risco, $SMR_i = \frac{Y_i}{E_i}$, e $\frac{\hat{\nu}}{\hat{\alpha}}$, a média estimada da distribuição *a priori* de θ_i. O valor observado SMR_i é encolhido na direcção da média populacional global - quando há um grande número observado de casos a estimativa está perto do correspondente SMR_i, caso contrário fica mais próxima da média global do risco.

Um exemplo da aplicação deste método aos dados do cancro do estômago com que temos vindo a trabalhar mas separados por género pode ser consultada em Natário (2005) onde outros métodos de modelação mais sofisticados são também apresentados.

Na Figura 3.4 apresentamos um mapa coropleto das estimativas de Clayton e Kaldor para a razão de mortalidade padronizada do cancro de estômago com que temos vindo a trabalhar. Para melhor apreciar os efeitos deste método, apresenta-se também um coropleto dos resíduos do modelo.



Estimativas CK Resíduos

Figura 3.4 *Estimativas de Clayton e Kaldor para a razão de mortalidade padronizada (esquerda) e resíduos do modelo (direita) .*

Como pode verificar-se pelo mapa da esquerda o modelo alisou bastante as observações do risco relativo. Grande parte do mapa apresenta agora tons associados a valores em torno de 1, risco que corresponde ao da média do país. Sobressaem ainda um conjunto de regiões a norte com valores bastante elevados e um conjunto menos definido a sul com valores também menos elevados que podem indiciar que a associação espacial se mantém mesmo para os valores alisados.

Nota-se pelo mapa dos resíduos à direita que tanto os valores ele-

vados como os valores baixos foram "puxados" para a média, mas não o suficiente para fazer desaparecer os aglomerados fortemente marcados dos dados originais.

Estes resultados indicam que a presença destes aglomerados não é casual e fazem suspeitar que deverá existir alguma razão, ou razões, para a sua presença. Eventuais variáveis explicativas poderão ser avaliadas através da criação de modelos que as incorporem e posterior análise dos resíduos com o intuito de verificar a permanência de associação espacial que poderá também vir a ser introduzida no modelo.

Chegamos assim à necessidade natural de modelar os dados referentes a áreas e, no que se segue, vamos apresentar alguns dos modelos mais comuns. Esses modelos fazem depender os estimadores de cada região apenas dos valores observados nas regiões vizinhas ao contrário do modelo que acabámos de apresentar onde, um pouco artificialmente, as estimativas dependiam de igual modo de todas as observações, mesmo as mais distantes.

3.4 Campos Aleatórios de Markov

Consideremos, como habitualmente, o domínio $\mathcal{D} \subseteq \mathbb{R}^2$, subdividido num conjunto de n blocos, notados simplesmente por $S = \{1, 2, \ldots, n\}$, sobre os quais está definido o processo $Y = \{Y_s : s \in S\}$, que interpretaremos como um vector n-dimensional, $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)'$. Designemos por E o espaço de estados do processo e por configuração de observações qualquer vector $(y_1, y_2, \ldots, y_n) \in E^n$.

A estrutura de dependência entre as variáveis deste campo aleatório discreto pode ser definida de duas formas: à custa da distribuição conjunta de $\boldsymbol{Y}$, $p(\boldsymbol{y}) = p(y_1, y_2, \ldots, y_n)$, modelando directamente a associação espacial pela matriz de covariâncias; através das distribuições

condicionais completas

$$p(y_i|y_j, j \neq i), \quad i = 1, 2, \ldots, n.$$

Note-se que o conhecimento da distribuição conjunta de $\boldsymbol{Y}$ permite a definição de todas as distribuições condicionais completas mas, como veremos no teorema que se segue, estas ficam sujeitas a condições de consistência muito exigentes. Essas condições são as que permitem que estas distribuições condicionais se ajustem de um modo não contraditório tornando possível a construção da verosimilhança para uma qualquer subcolecção finita de locais.
Para apresentarmos este importante resultado de Besag (1974), conhecido pelo Teorema da Factorização, temos primeiro que referir a condição designada por Hammersley and Clifford (1971) como **Condição de Positividade**

Definição 3.1. *Uma distribuição de probabilidade $p(\boldsymbol{y})$ diz-se que satisfaz a condição de positividade se, quaisquer que sejam $j \in S$, $y'_j \in E$ e $(y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_n) \in E^{n-1}$, se tem*

$$p_j(y'_j) = 0 \Longrightarrow p(y_1, \ldots, y_{j-1}, y'_j, y_{j+1}, \ldots, y_n) = 0$$

onde p_j representa a densidade marginal do bloco j.

A mesma definição pode escrever-se de forma recíproca:

$$p(y_j) > 0, \quad j = 1, \ldots, n \Longrightarrow p(y_1, \ldots, y_n) > 0, \tag{3.6}$$

ou seja, se $y_1, y_2, \ldots, y_n$ podem ocorrer individualmente nas posições $1, 2, \ldots, n$ então podem ocorrer conjuntamente.
Passemos ao Teorema da Factorização também conhecido por

Lema 3.2.1. **(Lema de Brook, 1964)**
Considere-se uma distribuição de probabilidade satisfazendo à condição de positividade. Para todos os $\boldsymbol{y}, \boldsymbol{y}' \in E^n$ com probabilidades

estritamente positivas, tem-se

$$\frac{p(\boldsymbol{y})}{p(\boldsymbol{y}')} = \frac{p(y_1|y_2,\ldots,y_n)}{p(y_1'|y_2,\ldots,y_n)}\frac{p(y_2|y_1',y_3,\ldots,y_n)}{p(y_2'|y_1',y_3,\ldots,y_n)}\cdots\frac{p(y_n|y_1',\ldots,y_{n-1}')}{p(y_n'|y_1',\ldots,y_{n-1}')}. \tag{3.7}$$

Demonstração:

O lema é fácil de demonstrar recursivamente começando com a igualdade simples de verificar,

$$p(\boldsymbol{y}) = \frac{p(y_n|y_1,\ldots,y_{n-1})}{p(y_n'|y_1,\ldots,y_{n-1})}p(y_1,\ldots,y_{n-1},y_n').$$

Usando a mesma técnica para $p(y_1,\ldots,y_{n-1},y_n')$ pode escrever-se

$$p(\boldsymbol{y}) = \frac{p(y_n|y_1,\ldots,y_{n-1})}{p(y_n'|y_1,\ldots,y_{n-1})}\frac{p(y_{n-1}|y_1,\ldots,y_{n-2},y_n')}{p(y_{n-1}'|y_1,\ldots,y_{n-2},y_n')}p(y_1,\ldots,y_{n-2},y_{n-1}',y_n').$$

A igualdade pretendida obtém-se prosseguindo de forma análoga para os restantes valores e tendo em conta a condição de positividade (3.6) e a positividade estrita de $p(\boldsymbol{y})$ e de $p(\boldsymbol{y}')$. ■

Como se vê, a natureza construtiva da fórmula (3.7) permite um cálculo simples da distribuição conjunta, mas também mostra como as condições de consistência para as condicionais completas são fortes. De facto, para as $n!$ ordenações diferentes dos n blocos, as correspondentes factorizações têm que dar o mesmo valor para o quociente $p(\boldsymbol{y})/p(\boldsymbol{y}')$, o que restringe consideravelmente as expressões funcionais das distribuições condicionais.

Embora o estudo de diferentes condições em que se verifica a compatibilidade das distribuições condicionais completas tenha merecido atenção (por exemplo, Arnold and Strauss, 1991), não vamos tratar desse ponto. Apresentamos, no entanto, um exemplo que mostra como situações de incompatibilidade são fáceis de encontrar.

Considerem-se as duas distribuições condicionais

$$Y_1|Y_2 \sim \mathcal{N}(\alpha_0 + \alpha_1 Y_2, \sigma^2) \qquad Y_2|Y_1 \sim \mathcal{N}(\beta_0 + \beta_1 Y_1^3, \sigma^2).$$

É fácil ver que

$$E[Y_1] = E[E[Y_1|Y_2]] = \alpha_0 + \alpha_1 E[Y_2]$$

e

$$E[Y_2] = E[E[Y_2|Y_1]] = \beta_0 + \beta_1 E[Y_1^3]$$

o que conduz a

$$E[Y_1] = \alpha_0 + \alpha_1(\beta_0 + \beta_1 E[Y_1^3]) = (\alpha_0 + \alpha_1\beta_0) + \alpha_1\beta_1 E[Y_1^3],$$

o que só é verdade em situações triviais.

Um segundo problema fica patente do enunciado do lema. Este permite calcular o valor relativo da distribuição conjunta em duas configurações fixas, $p(\boldsymbol{y})/p(\boldsymbol{y}')$, mas a distribuição conjunta $p(\boldsymbol{y})$, $\boldsymbol{y} \in E^n$, só fica especificada pelas distribuições condicionais completas a menos de uma constante de proporcionalidade $p(\boldsymbol{y}')$ que, em termos algorítmicos, só poderá ser conhecida se a distribuição for integrável.

No exemplo que se segue a distribuição conjunta não é integrável. Sejam $Y_1|Y_2 \sim \mathcal{N}(Y_2, 1)$ e $Y_2|Y_1 \sim \mathcal{N}(Y_1, 1)$. Aplicando o lema de Brook obtém-se

$$p(y_1, y_2) = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_1-y_2)^2}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_1'-y_2)^2}} \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_2-y_1')^2}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_1'-y_2')^2}} p(y_1', y_2') \propto e^{-(y_1-y_2)^2}$$

que facilmente se vê ser imprópria.

Passemos agora a definir o que se entende por campo aleatório de Markov.

Quando se está a trabalhar com um processo estocástico de parâmetro discreto unidimensional $\{Y_n, n \geq 0\}$ com valores em E, a propriedade de Markov,

$$P[Y_{i_n+m} \in I | Y_{i_1} = j_1, Y_{i_2} = j_2, \ldots, Y_{i_n} = j_n] = P[Y_{i_n+m} \in I | Y_{i_n} = j_n]$$

válida quaisquer que sejam $0 \leq i_1 \leq \cdots \leq i_n$, $m \geq 1$, $(j_1, \ldots, j_n) \in E^n$, e o intervalo $I \subset E$, pode também ser definida da seguinte forma:

Para $n \geq 1$, Y_n é independente de $(Y_k : k \notin \{n-1, n, n+1\})$, dados (Y_{n-1}, Y_{n+1}).

Outra forma de escrever este resultado é

$$Y_n \perp \Big(Y_k : k \notin \big(\{n\} \bigcup \partial_n\big)\Big) \;\Big|\; \Big(Y_k : k \in \partial_n\Big)$$

onde ∂_n é o conjunto dos vizinhos do instante n, ou seja, $\{n-1, n+1\}$.

Esta é a definição da propriedade de Markov mais conveniente para ser estendida a processos em $\mathbb{R}^2$. É o que faremos de seguida com a introdução da noção de **Campo Aleatório de Markov** que designaremos pelas iniciais de *Markov Random Field* (MRF).
Para $A \subset S$, seja $\boldsymbol{Y}(A) = \{Y_s : s \in A\}$. Com esta notação pode escrever-se $\boldsymbol{Y} = \{\boldsymbol{Y}(A), \boldsymbol{Y}(S-A)\}$. Temos então,

Definição 3.2. *Um campo aleatório é um MRF se para todo o $s \in S$ se verifica que*

$$Y_s \perp \boldsymbol{Y}\big(S - (\{s\} \cup \partial_s)\big) \;\Big|\; \boldsymbol{Y}(\partial_s).$$

Como se vê, esta definição estende de forma directa a definição de processo de Markov a tempo discreto que demos inicialmente, embora com imposições adicionais sobre a forma como a estrutura de vizinhanças deve ser definida.

Designemos por $\sim$ a relação de vizinhança entre blocos. Então,

(a) $\partial_s = \{t : t \sim s\}$,

(b) $s \notin \partial_s$,

(c) $s \in \partial_t \Longrightarrow t \in \partial_s$.

Para um MRF a especificação das distribuições condicionais completas pode ser feita exclusivamente à custa dos vizinhos de cada bloco.

De facto, chamamos **característica local** do MRF em s à distribuição Y_s condicionada apenas pelos seus vizinhos

$$p(y_s|y_{s'}, s' \in \partial_s) = P[Y_s = y_s|\boldsymbol{Y}(\partial_s) = \boldsymbol{y}(\partial_s)]$$

e tem-se

$$p(y_s|y_{s'}, s' \neq s) = p(y_s|y_{s'}, s' \in \partial_s).$$

À família das características locais $\{p(y_s|y_{s'}, s' \in \partial_s) : s \in S\}$ chama-se a **especificação local** do MRF.

De acordo com o Teorema da Factorização podemos então dizer que a especificação local tem que obedecer a certas condições de consistência para que se possa definir a distribuição conjunta dum campo aleatório que lhe esteja associado. São essas condições, também conhecidas pelas condições para a especificação local, que conduzem a um MFR viável, que vamos passar a estudar.

3.4.1 Cliques, Potênciais e Distribuições de Gibbs

Um papel determinante na construção de MRFs viáveis vai ser desempenhado pelas distribuições de Gibbs que apareceram na literatura da física por volta de 1902 pela mão deste investigador. À parte uma constante, a expressão geral das distribuições que Gibbs (1902) introduziu é dada por

$$p(\boldsymbol{y}) \propto exp(-\frac{1}{T}\mathcal{E}(\boldsymbol{y})) \tag{3.8}$$

em que T é um parâmetro conhecido por **temperatura** e $\mathcal{E}(\boldsymbol{y})$ por **energia** da configuração $\boldsymbol{y}$. Este tipo de distribuições são interessantes para os físicos se a energia se pode exprimir em termos de funções, ditas funções potencial, que descrevem as interacções locais.

Suponhamos que se encontra definida uma estrutura de vizinhanças sobre o conjunto S das n localizações a que se referem as configurações $\boldsymbol{y}$. Tem-se então,

Definição 3.3. *Um subconjunto $C \subseteq S$ chama-se* **clique** *de S se consiste num único elemento, ou num conjunto de elementos que são todos vizinhos uns dos outros. A cardinalidade de uma clique designa-se por* **ordem**. *Uma clique é* **maximal** *se a introdução de um outro qualquer elemento a impede de ser uma clique.*

Definição 3.4. *Seja C_k um subconjunto qualquer de S com k elementos, $k = 1, \ldots, n$. Uma função $\phi_{C_k}^{(k)} : E^n \mapsto \mathbb{R}$, diz-se uma* **função potencial de ordem** k **sobre** C_k *se verifica as seguintes propriedades.*

(i) $\phi_{C_k}^{(k)} \equiv 0$ *se C_k não é uma clique de ordem k.*

(ii) *Para quaisquer duas configurações $\boldsymbol{y}$, $\boldsymbol{y}'$ e todo o conjunto C_k,*

$$\boldsymbol{y}(C_k) = \boldsymbol{y}'(C_k) \Longrightarrow \phi_{C_k}^{(k)}(\boldsymbol{y}) = \phi_{C_k}^{(k)}(\boldsymbol{y}').$$

Quando a função energia se pode escrever sob a forma

$$\mathcal{E}(\boldsymbol{y}) = \sum_k \sum_{C_k} \phi_{C_k}^{(k)}(\boldsymbol{y}), \tag{3.9}$$

i.e., como soma de potenciais em cliques, diz-se que a energia **deriva de um potencial**.

São as distribuições (3.8) em que a energia deriva de um potencial que se chamam **distribuições de Gibbs**.

Vejamos como este tipo de distribuições surge na teoria dos MRFs.

Suponhamos, sem perda de generalidade que, $\mathbf{0} \in S_p$, com S_p suporte da distribuição de $\boldsymbol{Y}$, e designemos por **função potencial negativo** a seguinte transformação de $p(\cdot)$,

$$Q(\boldsymbol{y}) = \log \frac{p(\boldsymbol{y})}{p(\mathbf{0})}. \tag{3.10}$$

É fácil verificar que o conhecimento de $Q(\boldsymbol{y})$ é equivalente ao conhe-

cimento de $p(\boldsymbol{y})$ visto que de (3.10) se tem

$$\exp(Q(\boldsymbol{y})) = \frac{p(\boldsymbol{y})}{p(\mathbf{0})} \Rightarrow \sum_{\boldsymbol{y}\in S_p} p(\boldsymbol{y}) = p(\mathbf{0}) \sum_{\boldsymbol{y}\in S_p} \exp(Q(\boldsymbol{y}))$$
$$\Leftrightarrow p(\mathbf{0}) = \frac{1}{\sum_{\boldsymbol{y}\in S_p} \exp(Q(\boldsymbol{y}))},$$

ou ainda,

$$p(\boldsymbol{y}) = \frac{\exp(Q(\boldsymbol{y}))}{\sum_{\boldsymbol{y}\in S_p} \exp(Q(\boldsymbol{y}))} \tag{3.11}$$

em que a constante normalizadora $\sum_{\boldsymbol{y}\in S_p} \exp(Q(\boldsymbol{y}))$ é designada por **função partição**.

A expressão (3.11) mostra também que $p(\boldsymbol{y})$ se pode escrever sob a forma de uma distribuição do tipo (3.8) em que o papel da energia $\mathcal{E}(\boldsymbol{y})$ é desempenhado por $-Q(\boldsymbol{y})$ e a temperatura é unitária. A função partição desempenha o papel de constante normalizadora.

Os teoremas que vamos apresentar seguidamente estabelecem que todos os campos aleatórios de Markov têm a função potencial negativo do tipo (3.9) e que são os únicos com essa propriedade.

Teorema 3.3. **Geman and Geman(1984)**.

Se um campo aleatório tem uma distribuição de Gibbs, ou seja, se

$$p(\boldsymbol{y}) \propto \exp\big(-\sum_k \sum_{C_k} \phi_{C_k}^{(k)}(\boldsymbol{y})\big),$$

então o campo aleatório é um MRF com especificação local dada por

$$p(y_i|y_j \in \delta_i) = \frac{\exp\big(-\sum_k \sum_{C_k \ni i} \phi_{C_k}^{(k)}\big)}{\sum_y \exp\big(-\sum_k \sum_{C_k \ni i} \phi_{C_k}^{(k)}\big)}.$$

Teorema 3.4. **Hammersley and Clifford (1971)**[2].

Seja $p(\boldsymbol{y})$ a distribuição de um MRF que satisfaz a condição de positividade então $p(\boldsymbol{y})$ é uma distribuição de Gibbs.

[2]Trata-se de um manuscrito que nunca foi publicado. A primeira publicação deste resultado pertence a Besag(1974).

Vejamos alguns exemplos de aplicação destes teoremas:

- Se a distribuição de um campo aleatório usa apenas cliques de ordem 1 então o campo é constituído por variáveis aleatórias independentes.

- Se um campo aleatório tem distribuição conjunta dada por

$$p(y_1, \ldots, y_n) \propto \exp\{-\frac{1}{2}\sum_{ij}(y_i - y_j)^2 I_{(i\sim j)}\}, \tag{3.12}$$

então trata-se de um MRF com especificação local

$$p(y_i|y_j, j \in \partial_i) \sim \mathcal{N}\Big(\sum_{j\in\delta_i} y_j/m_i, \tau^2/m_i\Big),$$

em que m_i é o número de vizinhos da célula i.

Para terminar esta secção queremos fazer notar que estes resultados não vão ter apenas importância no reconhecimento de campos de Markov. De facto, eles vão permitir que, dada a especificação local de um MRF, se possam simular através de amostragem de Gibbs realizações da distribuição conjunta visto sabermos que esta distribuição existe e é única.

3.5 Modelação de processos referentes a áreas

Nesta secção vamos apresentar alguns modelos para a distribuição conjunta dos processos referentes a áreas para dados contínuos. Vamos apenas tratar o caso dessa distribuição ser gaussiana por existirem muitas aplicações cujos dados, ou transformações destes, podem ser modeladas através dela. Para uma introdução a modelos do mesmo tipo aplicados a dados não gaussianos sugere-se a leitura de Cressie (1993).

Apresentaremos apenas os dois modelos mais usualmente utilizados: aquele em que a estrutura de dependência é definida condicionalmente, modelos auto-regressivos condicionais CAR, e aqueles em que a estrutura de dependência é introduzida através da distribuição conjunta, modelos auto-regressivos simultâneos SAR.

Estes modelos, designados por Besag (1974) de **autonormais**, fazem parte de uma classe mais vasta de modelos, também por ele designados de **automodelos**, que são definidos por uma função potencial negativo do tipo

$$Q(\boldsymbol{y}) = \sum_i y_i G_i(y_i) + \sum_i \sum_j \beta_{ij} y_i y_j,$$

em que β_{ij} é nulo a menos que i e j sejam vizinhos.

3.5.1 Modelos auto-regressivos condicionais (CAR)

Os modelos CAR foram apresentados pela primeira vez por Besag no trabalho de 1974 que já citámos e, como o nome indica, baseiam-se na especificação local de um MRF.

Suponhamos o seguinte conjunto de condicionais completas

$$Y_i | Y_j, j \neq i \Leftrightarrow Y_i \big| \boldsymbol{Y}(S - \{i\}) \sim \mathcal{N}\left(\sum_j c_{ij} y_j, \tau_i^2 \right), \quad (3.13)$$
$$i = 1, \ldots, n.$$

Através do Lema de Brook pode encontrar-se a distribuição conjunta de $\boldsymbol{Y}$, que é dada por,

$$p(y_1, \ldots, y_n) \propto \exp\Big\{ -\frac{1}{2} \boldsymbol{y}' \underbrace{D^{-1}(I - C)}_{\Sigma_y^{-1}} \boldsymbol{y} \Big\} \quad (3.14)$$

em que $C = (c_{ij})$ e D é uma matriz diagonal com $d_{ii} = \tau_i^2$.

A expressão (3.14) sugere imediatamente que $\boldsymbol{Y}$ tem uma distribuição gaussiana multivariada com valor médio nulo e matriz de covariância $\Sigma_y = (I - C)^{-1}D$ e, como tal, esta matriz deverá ser simétrica. Para assegurar que Σ_y^{-1} é simétrica teremos que exigir

$$\Sigma_{ij} = \frac{c_{ij}}{\tau_i^2} = \Sigma_{ji} = \frac{c_{ji}}{\tau_j^2}, \quad i \neq j \tag{3.15}$$

o que implica [3]:

$$\begin{cases} \text{se } C \text{ simétrica} \Longrightarrow \tau_i = \tau_j = \tau, \quad i, j = 1, \ldots, n \\ \text{se } C \text{ não simétrica} \Longrightarrow \exists (i,j), i \neq j : \tau_i \neq \tau_j. \end{cases}$$

Consideremos de agora em diante o caso em que W é uma matriz de vizinhanças simétrica.

Um modelo particular em que a condição (3.15) é verificada obtém-se fazendo

$$c_{ij} = \frac{w_{ij}}{w_i^+} \quad \text{e} \quad \tau_i^2 = \frac{\tau^2}{w_i^+},$$

com $w_i^+ = \sum_j w_{ij}$ representando, como habitualmente, o número de vizinhos.

Neste caso, as distribuições condicionais tomam a forma,

$$Y_i \big| \boldsymbol{Y}(S - \{i\}) \quad \sim \quad \mathcal{N}\left(\sum_j \frac{w_{ij}}{w_i^+} y_j, \frac{\tau^2}{w_i^+} \right) \qquad i = 1, \ldots, n, \tag{3.16}$$

e a distribuição conjunta (3.14), dada pelo lema de Brook, tem agora a expressão mais simples

$$p(y_1, \ldots, y_n) \propto \exp\{-\frac{1}{2\tau^2} \boldsymbol{y}' \underbrace{(D_W - W)}_{\Sigma_y^{-1}} \boldsymbol{y}\} \tag{3.17}$$

onde D_W é uma matriz diagonal com $d_{Wii} = w_i^+$.

[3] A matriz de pesos W é suposta irredutível.

Este modelo que parece ser interessante tem contudo um problema: A matriz Σ_y^{-1} é singular, ou seja, $(D_W - W)\mathbf{1} = \mathbf{0}$, como é fácil verificar,

$$\left[\begin{pmatrix} w_1^+ & & \\ & \ddots & \\ & & w_n^+ \end{pmatrix} - \begin{pmatrix} w_{11} & \dots & w_{1n} \\ & \ddots & \\ w_{n1} & \dots & w_{nn} \end{pmatrix}\right] \times \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} w_1^+ - w_{11} - \cdots - w_{1n} \\ \vdots \\ w_1^+ - w_{n1} - \cdots - w_{nn} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

A consequência deste facto é que a distribuição conjunta é não integrável apesar das condicionais completas serem todas próprias [4]. Isto é fácil de verificar sabendo que, depois de alguns cálculos, se pode reescrever a expressão (3.17) como,

$$p(y_1, \dots, y_n) \propto \exp\{-\frac{1}{2\tau^2} \sum_{i \neq j} w_{ij}(y_i - y_j)^2\}, \tag{3.18}$$

em que a distribuição conjunta fica especificada por diferenças de pares, tal como acontecia com a distribuição menos geral introduzida em (3.12) como exemplo de um campo aleatório de Markov.

De facto, adicionando uma constante a todas as variáveis o valor da distribuição (3.18) não fica afectado. Tal significa que, desde que se mantenha a relação entre valores das variáveis, estes podem crescer sem limite mantendo a probabilidade de ocorrência, o que faz com que o seu integral seja infinito. Isto também quer dizer que as variáveis não ficam identificadas sendo necessário, para que passem a

[4] Neste caso, sabe-se da teoria dos grafos que, se a matriz W é irredutível, a característica da matriz $\Sigma^{-1} = D_W - W$ é igual a $n - 1$. Assim para tornar a distribuição própria é apenas necessário juntar uma condição extra sobre as variáveis.

sê-lo, impor-lhes uma condição suplementar. Normalmente a condição utilizada é

$$\sum_{i=1}^{n} Y_i = 0. \tag{3.19}$$

Os modelos que são especificados pelas distribuições condicionais (3.16), (ou pela conjunta(3.18)) são conhecidos por modelos **intrinsecamente auto-regressivos** e, pelo que se acabou de verificar, não podem ser aplicados a dados observados. Efectivamente, não existem mecanismos geradores de dados reais que sejam impróprios e não podemos impor uma condição do tipo (3.19) a observações.

Como consequência, a aplicação dos modelos intrinsecamente auto-regressivos fica restrita a casos que obedeçam naturalmente a esta condição, como acontece com os resíduos de um modelo de regressão, ou a outros onde se possa impô-la. É o que acontece no contexto da modelação hierárquica em que a correlação espacial é incorporada no modelo através de um vector de efeitos aleatórios $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)'$. Pode ter-se, por exemplo,

$$Y_i \overset{ind}{\sim} \mathcal{N}(\phi_i, \sigma^2), \quad i = 1, \ldots, n,$$

e usar como priori para os efeitos aleatórios um modelo auto-regressivo intrínseco

$$\phi_i | \boldsymbol{\phi}(S - \{i\}) \sim \mathcal{N}\left(\sum_j \phi / w_i^+, \tau^2 / w_i^+ \right).$$

Uma vez que os modelos intrinsecamente auto-regressivos tem as limitações que acabámos de verificar, construiu-se, a partir do tipo de especificação (3.17) do modelo particular que temos vindo a estudar, um outro modelo aparentemente mais geral que envolve um parâmetro adicional.

A ideia é redefinir a matriz inversa da covariância como

$$\Sigma_y^{-1} = D_W - \rho W$$

e escolher o novo parâmetro ρ de forma a que a matriz Σ_y^{-1}seja uma matriz invertível.

Isto consegue-se se ρ for escolhido dentro do intervalo $(\frac{1}{\lambda_{(1)}}, \frac{1}{\lambda_{(n)}})$, em que se designam por $\lambda_{(i)},\ i = 1, \ldots, n$, os valores próprios ordenados crescentemente da matriz $D_W^{-1/2} W D_W^{-1/2}$.

Note-se que os valores próprios da matriz anterior são tais que

$$\sum_{i=1}^{n} \lambda_{(i)} = \text{tr } (D_W^{-1/2} W D_W^{-1/2}) = 0$$

o que mostra a existência de valores próprios positivos e negativos [5] e, consequentemente, a inclusão do valor zero no intervalo de valores admissíveis para ρ.

Note-se, precisamente, que o caso $\rho = 0$ configura a situação limite em que as variáveis são independentes.

Uma vez que as distribuições condicionais associadas a

$$p(y_1, \ldots, y_n) \propto \exp\{-\frac{1}{2\tau^2} \boldsymbol{y}'(D_W - \rho W)\boldsymbol{y}\} \tag{3.20}$$

são

$$Y_i | \boldsymbol{Y}(S - \{i\}) \ \sim \mathcal{N}(\rho \sum_j \frac{w_{ij}}{w_i^+} y_j, \frac{\tau^2}{w_i^+}) \qquad i = 1, \ldots, n, \tag{3.21}$$

[5] Dada uma matriz A e uma matriz invertível P, a matriz $P'AP$ diz-se congruente com A. Matrizes congruentes com matrizes simétricas são simétricas.
Para matrizes com entradas em $\mathbb{R}$ verificam-se algumas propriedades especiais:
Uma matriz simétrica é sempre congruente com uma matriz diagonal.
Segundo o teorema de Sylvester (lei da inércia), todas as matrizes congruentes com uma matriz simétrica têm o mesmo número de valores próprios positivos e negativos (e portanto nulos).
Note que W é uma matriz simétrica congruente com $D_W^{-1/2} W D_W^{-1/2}$.

se $\rho = 0$, os Y_i passam a ser variáveis independentes com distribuição

$$Y_i \sim \mathcal{N}\left(0, \frac{\tau^2}{w_i^+}\right).$$

As vantagens deste modelo parecem ser imbatíveis. Excluindo o caso $\rho = 1$ que corresponde ao modelo intrinsecamente auto-regressivo, com uma escolha adequada de ρ esta especificação conduz a matrizes Σ^{-1} invertíveis e, consequentemente, a distribuições conjuntas próprias. Além disso, o caso extremo de independência das variáveis do processo obtém-se com $\rho = 0$.

Infelizmente, embora aparentemente este modelo cubra uma gama muito vasta de situações, tem a grande desvantagem de possibilitar uma pequena banda de hipóteses para a associação entre as variáveis componentes do processo.

Para verificarmos que assim é, vamos olhar de novo para a expressão (3.21). Note-se que o valor médio da distribuição condicional de Y_i não é a média dos valores dos seus vizinhos mas sim uma parte dela

$$\rho \sum_{i=1}^{n} \frac{w_{ij}}{w_i^+} y_j.$$

Desta forma, ρ pode ser encarado como uma medida da forma como Y_i é, em média, afectado pelos valores dos seus vizinhos. Seria, por conseguinte, interessante tentar relacionar este parâmetro com uma medida de associação espacial, como por exemplo o I de Moran ou o c de Geary, e verificar se os valores possíveis para ρ produzem modelos cobrindo a extensão das possibilidades de associação espacial.

Ora, vários estudos de simulação (ver por exemplo, Banerjee and al. 2004) mostram que, mesmo para valores muito altos de ρ, como por exemplo 0.99, o valor de I não ultrapassa 0.5, (para $\rho = 0.9$, $I \in [0.2, 0.25]$), o que mostra como são limitadas as possibilidades deste modelo explicar razoavelmente padrões espaciais fortes e indica

que será o modelo com $\rho = 1$ o mais adequado para modelar essas situações. Vemo-nos, assim, obrigados a regressar à alternativa de utilizar um modelo com efeitos aleatórios, modelados a priori por um CAR impróprio , admissível desde que acompanhado da condição (3.19), obtendo uma posteriori com uma maior amplitude de padrões de associação espacial.

Voltemos ao modelo inicial (3.13) e escrevamo-lo sob a forma de um modelo auto-regressivo:

$$\boldsymbol{Y} = C\boldsymbol{Y} + \boldsymbol{\epsilon}$$

ou, equivalentemente, $(I - C)\boldsymbol{Y} = \boldsymbol{\epsilon}$.

Comecemos por verificar as propriedades de $\boldsymbol{\epsilon}$.

A atribuição de uma distribuição a $\boldsymbol{Y}$ induz uma distribuição para $\boldsymbol{\epsilon}$ e, assim, uma vez que se sabe que a este modelo CAR corresponde a uma distribuição conjunta gaussiana

$$\boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, (I - C)^{-1}D), \tag{3.22}$$

vem

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, D(I - C)').$$

De facto, supondo que $(I - C)$ é não singular,

$$\Sigma_\epsilon = \underbrace{(I - C)(I - C)^{-1}}_{I} D(I - C)' = D(I - C)'$$

o que significa que as inovações não são independentes entre si como acontece habitualmente nos modelos de regressão. No entanto, também se pode verificar que $\mathrm{Cov}(\boldsymbol{\epsilon}, \boldsymbol{Y}) = D$, matriz diagonal, o que significa que ϵ_i só depende de Y_i, ou seja o valor da variável num local Y_i só depende da inovação nesse local e é independente das inovações em todos os outros locais $\epsilon_j, j \neq i$.

Mais alguns pontos merecem uma referência especial.

- A variação em larga escala dos dados é normalmente integrada nos modelos CAR através de uma componente de regressão. Esta componente é adicionada à estrutura dos valores médios das observações. Assim, designando por X a matriz dos valores de p variáveis explicativas nos n blocos, e por $\boldsymbol{\beta}$ o vector dos respectivos parâmetros de regressão, a distribuição conjunta é dada por

$$\boldsymbol{Y} \sim \mathcal{N}(X\boldsymbol{\beta}, (I-C)^{-1}D).$$

Assim sendo, as distribuições condicionais são dadas por

$$(Y_i - \sum_{k=1}^{p} x_{ik}\beta_k) \;\Big|\; (Y_j - \sum_{k=1}^{p} x_{jk}\beta_k),\ i \neq j \sim \mathcal{N}\left(\sum_j \frac{w_{ij}}{w_i^+} y_j, \frac{\tau^2}{w_i^+}\right)$$
$$i = 1, \ldots, n,$$

ou equivalentemente,

$$(I-C)(Y - X\boldsymbol{\beta}) = \boldsymbol{\epsilon}.$$

Repare-se que, condicionalmente em $\boldsymbol{\beta}$, a introdução desta componente não afectou a estrutura da associação que está subjacente ao modelo, só modificou a estrutura dos valores médios.

- Note-se que se pode usar um modelo CAR para fazer predição dum valor da variável de interesse num bloco não observado, Y_0. No entanto, ao contrário do que acontecia para os dados referentes a pontos, a predição de Y_0 não está univocamente definida, dependendo da forma como se especifica a distribuição condicional de Y_0 dado $\boldsymbol{Y}$.

Poderemos, por exemplo, usar a especificação

$$Y_0 \mid \boldsymbol{Y} \sim \mathcal{N}\left(\rho \sum_j \frac{w_{0j} y_j}{w_0^+}, \frac{\tau^2}{w_0^+}\right),$$

e, simplesmente, predizer Y_0 pelo valor médio desta distribuição, com ρ já calculado através do modelo CAR aplicado a $\boldsymbol{Y}$.

Note-se, no entanto, que esta especificação vai determinar uma distribuição conjunta de $(Y_0, Y_1, \ldots, Y_n)$ que não é a distribuição CAR que apareceria se especificássemos as distribuições condicionais completas de Y_i, $i = 0, 1, \ldots, n$, e usássemos o lema de Brook. De facto, no primeiro caso, ρ seria calculado usando a matriz de vizinhanças inicial com $n \times n$ elementos, enquanto que no segundo caso, ρ seria calculado com base numa matriz de vizinhanças $(n+1) \times (n+1)$.

3.5.2 Modelos auto-regressivos simultâneos (SAR)

Os modelos SAR são definidos por

$$Y_i = \sum_j s_{ij} Y_j + \epsilon_i \quad i = 1, \ldots, n$$

com $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$, inovações independentes com variância distinta.

Estamos perante um conjunto de n modelos auto-regressivos simultâneos, que podem também ser descritos pela equação vectorial

$$\boldsymbol{Y} = S\boldsymbol{Y} + \boldsymbol{\epsilon}, \tag{3.23}$$

em que S é a matriz de entradas s_{ij}.

Neste caso, a distribuição de $\boldsymbol{Y}$ vai ser induzida pela distribuição de $\boldsymbol{\epsilon}$. Tomando em consideração que

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \widetilde{D}),$$

onde $\widetilde{D}$ é uma matriz diagonal com $\widetilde{d}_{ii} = \sigma_i^2$, e que, de (3.23) se tem,

$$(I - S)\boldsymbol{Y} = \boldsymbol{\epsilon} \Leftrightarrow \boldsymbol{Y} = (I - S)^{-1}\boldsymbol{\epsilon},$$

se $(I - S)$ é invertível, obtém-se, aplicando a propriedade das transformações lineares da multinormal,

$$Y \sim \mathcal{N}\left(\mathbf{0}, (I - S)^{-1}\widetilde{D}((I - S)^{-1})'\right). \tag{3.24}$$

Temos também[6], $\mathrm{Cov}(\boldsymbol{\epsilon}, \boldsymbol{Y}) = \widetilde{D}(I-S)^{-1}$.

Se assumirmos que as inovações têm igual variância, $\widetilde{D} = \sigma^2 I$, a distribuição conjunta de $\boldsymbol{Y}$ simplifica-se para

$$Y \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\big((I-S)'(I-S)\big)^{-1}\right), \tag{3.25}$$

em que, como já se disse, se supõe $(I-S)$ invertível.

As escolhas mais usuais para a matriz S são as seguintes:

1. $S = \rho W$ com W matriz de pesos.

Neste caso ρ é chamado **parâmetro de auto-regressão espacial** e tem-se,

$$Y_i = \rho \sum_j w_j Y_j + \epsilon_i \tag{3.26}$$

Para que $(I - \rho W)$ seja não singular, teremos que exigir, tal como deduzimos para os modelos CAR, que

$$\rho \in \left(\frac{1}{\lambda_{(1)}}, \frac{1}{\lambda_{(n)}}\right) \tag{3.27}$$

onde agora $\lambda_{(1)} \leq \cdots \leq \lambda_{(n)}$ são os valores próprios ordenados de W.

2. $S = \alpha\widetilde{W}$ com $\widetilde{W}$ matriz de pesos estandardizada por linhas, i.e., $\widetilde{w}_{ij} = wij/w_i^+$.

Neste caso, α é designado por **parâmetro de autocorrelação espacial** e tem-se,

$$Y_i = \alpha \sum_j Y_j \frac{w_{ij}}{w_i^+} + \epsilon_i.$$

A condição sobre α para que $(I - \alpha\widetilde{W})$ seja não singular é bastante simples e explica a designação deste parâmetro. Embora a matriz

[6]Note-se que esta situação é bem diversa da equivalente para os modelos CAR. Neste caso, embora as inovações sejam independentes entre si , o valor da variável num certo local pode depender das inovações noutros locais.

$\widetilde{W}$ não seja necessariamente simétrica, é estocástica por linhas e, portanto, todos os seus valores próprios terão módulo inferior a 1, i.e. $\max|\lambda_i| \leq 1$. Isto significa que, se tomarmos $\alpha \in (-1, 1)$, teremos uma condição do tipo (3.27) automaticamente verificada.

Os modelos SAR são normalmente introduzidos no contexto da regressão, ou seja, são os resíduos de uma regressão linear que explica as variações de grande escala, $\boldsymbol{U} = \boldsymbol{Y} - X\boldsymbol{\beta}$, que são supostos seguir um modelo SAR, e não as observações.

Usando a mesma notação matricial que em (3.23) obtém-se,

$$\begin{aligned}
\boldsymbol{U} &= S\boldsymbol{U} + \boldsymbol{\epsilon} \\
\boldsymbol{Y} - X\boldsymbol{\beta} &= S(\boldsymbol{Y} - X\boldsymbol{\beta}) + \boldsymbol{\epsilon} \\
\boldsymbol{Y} &= \underbrace{S\boldsymbol{Y}}_{(*)} + \underbrace{(I - S)X\boldsymbol{\beta}}_{(**)} + \boldsymbol{\epsilon} && (3.28) \\
&= S\boldsymbol{Y} + X\boldsymbol{\beta} - SX\boldsymbol{\beta} + \boldsymbol{\epsilon}. && (3.29)
\end{aligned}$$

Esta última expressão mostra que $\boldsymbol{Y}$ é dado como uma combinação de uma componente espacial $(*)$ que incorpora uma ponderação dos vizinhos e uma componente de regressão linear $(**)$. Esta última componente pode ainda ser dividida numa componente de regressão linear usual $X\boldsymbol{\beta}$ e numa outra componente de regressão linear sobre a matriz transformada SX. Se a matriz S é nula então o modelo é uma regressão linear comum.

Note-se que, usando $S = \alpha\widetilde{W}$ no modelo anterior, se obtém,

$$\begin{aligned}
\boldsymbol{Y} &= \alpha\widetilde{W}\boldsymbol{Y} + X\boldsymbol{\beta} - \alpha\widetilde{W}X\boldsymbol{\beta} + \boldsymbol{\epsilon} \\
&= \alpha\widetilde{W}\boldsymbol{Y} + X\boldsymbol{\beta} - \widetilde{W}X(\alpha\boldsymbol{\beta}) + \boldsymbol{\epsilon}
\end{aligned}$$

onde, além da componente de regressão usual nas covariáveis, se tem uma componente de regressão na média das covariáveis dos locais vizinhos.

Um modelo um pouco mais geral que este, foi criado por Durbin (1970). Ele permite que a componente de regressão na média dos vizinhos das covariáveis tenha um coeficiente independente do coeficiente da regressão nas covariáveis

$$\boldsymbol{Y} = \alpha\widetilde{W}\boldsymbol{Y} + X\boldsymbol{\beta} + \widetilde{W}X\boldsymbol{\alpha}_1 + \boldsymbol{\epsilon}$$

Os dois modelos serão iguais se $\boldsymbol{\alpha}_1 = -\alpha\boldsymbol{\beta}$.

3.5.3 Comparação dos modelos CAR e SAR

Foi Brook (1964) que fez a distinção entre a especificação CAR e a SAR para modelos espaciais. Neste texto só se trataram modelos CAR e SAR gaussianos e será exclusivamente a eles que nos reportamos. Uma comparação das expressões das distribuições conjuntas (3.22) e (3.24) mostra que os dois modelos são iguais desde que as matrizes de covariância sejam idênticas, isto é

$$(I - C)^{-1}D = (I - S)^{-1}\widetilde{D}\left((I - S)^{-1}\right)'$$

ou, no caso de $D = \tau^2 I = \sigma^2 I = \widetilde{D}$,

$$(I - C)^{-1} = (I - S)^{-1}\left((I - S)^{-1}\right)'$$

em que $\widetilde{D}$ refere a matriz do modelo SAR.
Cressie(1993) demonstrou que todo o modelo SAR é um modelo CAR[7] mas deu um contra-exemplo da afirmação contrária.
De facto, o modelo CAR alarga o comportamento das inovações, que no modelo SAR são necessariamente independentes, permitindo a existência de dependência entre inovações de locais diferentes.

[7] Basta fazer $C = S + S' - S'S$.

3.5.4 Aplicações

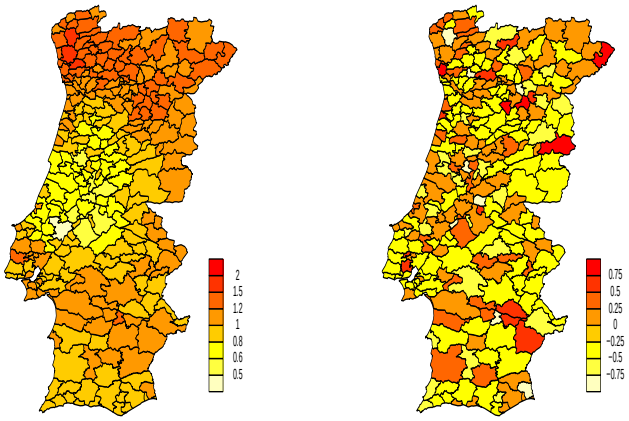
As figuras seguintes mostram os mapas coropletos das estimativas, produzidas pelos modelos CAR e SAR, da razão de mortalidade padronizada para o cancro de estômago que temos vindo a tratar. Uma vez que este risco relativo não tem distribuição gaussiana, utilizou-se para fazer o ajustamento dos modelos, a transformação dos dados, $Z_i = \log(Y_i + 0.5)/E_i$, que provou ser suficientemente para os simetrizar. Depois das estimativas calculadas aplicou-se a transformação inversa para que os mapas coropletos se pudessem comparar com os apresentados anteriormente. Os parâmetros de ambos os modelos foram estimados pelo método de máxima verosimilhança usando a matriz de vizinhanças por adjacência como matriz de pesos.

Na Figura 3.5 apresenta-se o mapa das estimativas do modelo CAR e dos correspondentes resíduos.

O valor estimado do parâmetro de auto-regressão ρ, referido em (3.20), é 0.157. Por ser significativamente diferente de zero indica a presença de associação espacial. O máximo do logaritmo da verosimilhança alcançado foi -57.748.

O mapa das estimativas do modelo, à esquerda, mostra um padrão semelhante ao apresentado pelos mapas dos dados, Figura 3.1, e das estimativas de Clayton e Kaldor, Figura 3.4, designadas de agora em diante por estimaticas CK.

São claros dois aglomerados de valores elevados, um a norte, mais pronunciado, e um outro a sul, separados por uma zona central de valores baixos. A transição entre estes aglomerados parece, no entanto, ser feita de forma mais suave que no caso das estimativas CK o que dá ao mapa um aspecto mais regular. Isto deve-se ao facto de, nos modelos auto-regressivos, a estimativa em cada bloco ser pri-



Estimativas CAR Resíduos

Figura 3.5 *Estimativas para a razão de mortalidade padronizada calculadas através do modelo CAR (esquerda) e resíduos do modelo (direita) .*

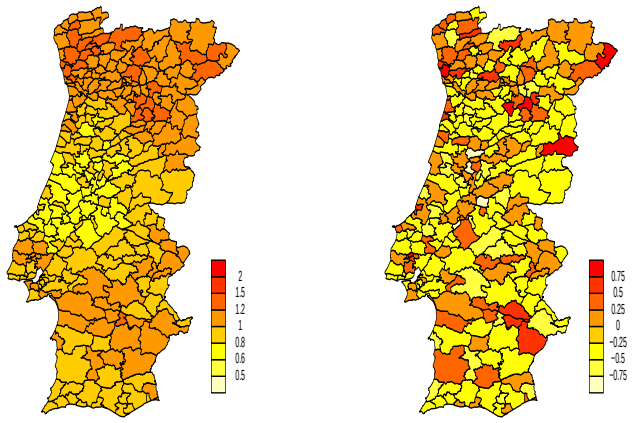
mordialmente influenciada pelos valores observados nos blocos mais próximos de si, os seus vizinhos por adjacência.

A observação do mapa dos resíduos, à direita, parece indicar, se exceptuarmos duas pequenas manchas a norte, uma de valores positivos e outra de valores negativos um pouco mais para sul, que os valores positivos e negativos se encontram espalhados ao acaso. Aliás a variância dos resíduos é baixa e a sua distribuição não parece muito enviesada o que indica que o modelo explica razoavelmente os dados.

Passemos agora à análise dos resultados da aplicação do outro modelo.

Na Figura 3.6 apresenta-se, à esquerda, o mapa das estimativas da razão de mortalidade padronizada usando o modelo SAR e, à direita,

o dos correspondentes resíduos.



Estimativas SAR Resíduos

Figura 3.6 *Estimativas para a razão de mortalidade padronizada calculadas através do modelo SAR (esquerda) e resíduos do modelo (direita) .*

O valor estimado do parâmetro de auto-regressão espacial ρ, referido em (3.26), é 0.112. Tal como no modelo CAR o seu valor é significativamente diferente de zero indicando a presença de associação espacial. O máximo do logaritmo da verosimilhança alcançado foi -63.567.

No geral, o aspecto do mapa das estimativas é semelhante ao dos outros modelos estudados, mas uma observação mais cuidadosa mostra que o modelo SAR alisa mais fortemente as observações aproximando-as ainda mais da média. O aspecto do mapa é também bastante regular pelas razões que apontámos anteriormente.

Os resíduos têm uma distribuição bastante equilibrada e, embora

apresentem, tal como no caso anterior, algumas pequenas manchas, a sua dimensão não parece contrariar a capacidade do modelo explicar o padrão espacial dos dados.

Evidentemente que os modelos que utilizámos neste trabalho dão uma pálida ideia do conjunto de opções que existem na modelação de dados referentes a áreas, tema que tem tido nos últimos tempos um grande desenvolvimento, especialmente no que diz respeito à utilização de modelos hierárquicos. Sugerimos a propósito os seguintes artigos Natário (2001), Natário e Carvalho(2002, 2005), Natário e Knorr-Held(2003), Held *et al.*.(2005) relacionados com este tipo de modelação, aplicados a dados de natureza epidemiológica.

A estimação de parâmetros para os modelos que aqui apresentámos foi deixada propositadamente de parte e sobre o assunto recomendamos a leitura de Cressie (1993) ou, mais especificamente para os modelos CAR, os artigos de Besag (1975, 1977) onde o autor introduz o método de estimação de máxima pseudoverosimilhança, que será apresentado no capítulo seguinte, e que está especialmente adaptado à estimação destes modelos.

Capítulo 4

Processos Pontuais Espaciais

4.1 Introdução e exemplos

A teoria dos processos pontuais surgiu com a necessidade de modelar como aleatória a localização de acontecimentos de interesse. As primeiras aplicações são da área da ecologia e das ciências florestais (Goodall, 1952, 1970; Pielou, 1977; Ripley, 1987). No entanto, estes processos viram ao longo dos anos diversificadas as suas aplicações: arqueologia (Hodder e Orton, 1976), cosmologia (Neyman e Scott, 1958), geografia (Cliff e Ord, 1981), sismologia (Ogata, 1989) e epidemiologia (Diggle, 1983).

Central à aplicação dos modelos de processos pontuais está a chamada **hipótese de aleatoriedade espacial completa**, que assume que o número de acontecimentos numa região planar segue uma distribuição de Poisson com média proporcional à área da região e ao número médio de acontecimentos por unidade de área (intensidade do processo) e que, condicionalmente ao número de acontecimentos na região ser n, as suas localizações formam aí uma amostra aleatória $n-$dimensional de uma distribuição uniforme.

A rejeição desta hipótese é um pré-requisito mínimo para qualquer tentativa de modelação de um padrão observado já que esta opera dividindo os padrões entre regulares e agregados (Mateu, 2004). Vários

testes baseados em distâncias e técnicas Monte Carlo foram desenvolvidos com este propósito (Diggle, 1983).

Na Figura 4.1 apresentam-se três padrões espaciais bastante diferentes, numa região quadrada (200 pontos em cada). Os pontos no gráfico (a) obedecem à hipótese de aleatoriedade completa atrás descrita, pelo que o padrão espacial correspondente não apresenta nenhuma estrutura aparente.

Os pontos em (b) encontram-se distribuídos regularmente na região. Estes padrões regulares surgem pela imposição de uma distância mínima entre acontecimentos (o que acontece, por exemplo, quando estudamos a localização de árvores, que competem por luz, e isso obriga a que mantenham uma distância mínima entre elas).

Finalmente o padrão observado em (c) corresponde a acontecimentos fortemente agregados, fruto de algum mecanismo de agregação ou apenas devido a variação ambiental.

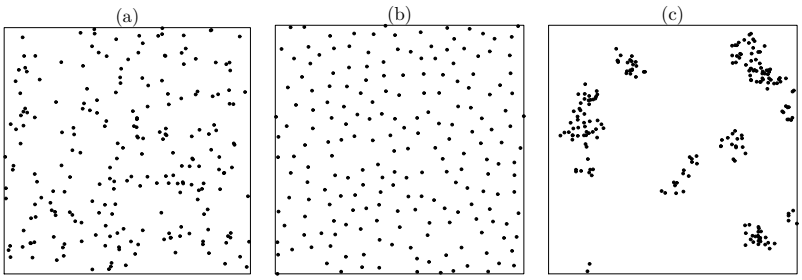


Figura 4.1 *Padrões simulados: (a) realização da hipótese de aleatoriedade completa; (b) padrão regular; (c) padrão agregado.*

Estabelecida a existência de um padrão espacial distinto do postulado pela hipótese de aleatoriedade espacial completa, surge a necessidade da modelação desses padrões, com a correspondente estimação

e testes aos valores dos parâmetros.

De início estes problemas restringiam-se a padrões pontuais com poucas observações, onde pressupostos de estacionaridade e isotropia eram essenciais, e os métodos assentavam essencialmente em abordagens não paramétricas, baseadas em estatísticas resumo. As referências clássicas mais importantes desta altura são Bartlett (1975); Cliff e Ord(1981); Ripley (1981; 1988); Diggle (1983); Upton e Fingleton(1985); Matérn(1986); Cressie (1991); Stoyan e Stoyan (1994); Stoyan *et al.*(1995).

Com o desenvolver dos recursos computacionais e, em particular com a aplicação de métodos de Monte Carlo e outros métodos de optimização numérica, o cenário alterou-se. Além disso, a inferência sobre os processos pontuais baseada na verosimilhança, cujo problema é a própria função de verosimilhança, quase sempre intratável matematicamente, pode agora ser feita recorrendo a um método de estimação baseado na chamada função de pseudoverosimilhança definida por Besag (1974, 1978).

Em resumo, a inferência que hoje em dia se faz recorre, frequentemente, a inferência baseada na verosimilhança, com o uso de modelos paramétricos flexíveis, muitas vezes incorporando covariáveis, liberta dos pressupostos restritivos atrás mencionados. De entre as referências modernas destacam-se van Lieshout (2000); Baddeley e Turner(2000); Diggle (2003); Møller e Waagepetersen (2003); Baddeley, Gregori, Mateu, Stoica e Stoyan (2006).

As aplicações que trataremos no decorrer deste capítulo são da área das ciências florestais e ambientais - localização de árvores e localização de incêndios em Portugal continental.

Os dois primeiros conjuntos de dados que apresentamos são clássicos da literatura dos processos pontuais, observados numa região quadrada. O primeiro são as localizações observadas de 65 mudas

de pinheiro negro Japonês num quadrado de 5.7 metros de lado (Numata, 1961) e o segundo é parte de um conjunto de dados extraído por Ripley (1977) de Strauss (1975), representando 62 localizações de rebentos de Pau-Brasil, num quadrado de lado 23 metros, aproximadamente - Figura 4.2.

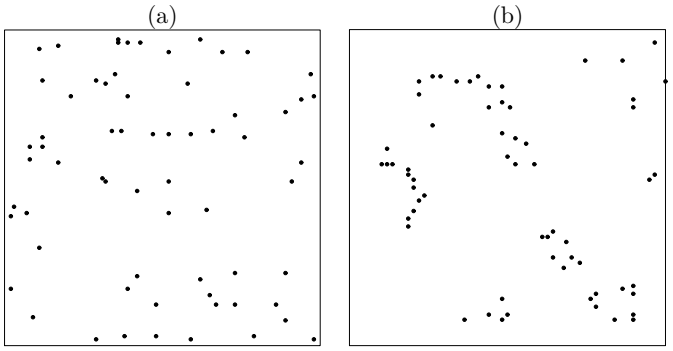


Figura 4.2 *Localizações de mudas de pinheiro negro Japonês (a) e de rebentos de Pau-Brasil (b).*

Enquanto que as mudas de pinheiro negro Japonês parecem localizar-se de forma 'completamente aleatória', os rebentos de Pau-Brasil aparecem claramente aglomerados. Esta última constatação não surpreende os biólogos, já que estes rebentos se tendem a aglomerar junto dos troncos do Pau-Brasil que lhes deram origem.

Os padrões espaciais podem ser mais complicados do que as figuras anteriores sugerem. A região de amostragem dos pontos pode ter uma forma arbitrária qualquer, em vez de ser um quadrado, como é o caso dos dados que se seguem.

No terceiro exemplo temos as localizações da ocorrência de todos os incêndios em Portugal continental, no ano de 1975 - Figura 4.3 (a).

Estes dados são um subconjunto de uma base de dados mais vasta que tem as localizações de todos os incêndios ocorridos nesta região, de 1975 a 2005 - dados gentilmente cedidos pelo Prof. José Miguel Pereira do Instituto Superior de Agronomia.

Temos ainda a informação da área ardida em cada fogo, que está também representada, para o ano de 1975, na Figura 4.3 (b) - sendo a área de cada circunferência centrada na localização de cada incêndio proporcional à referida área.

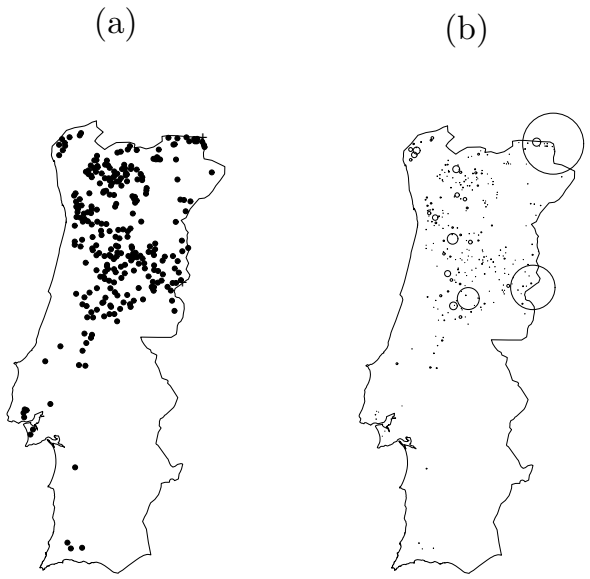


Figura 4.3 *Localizações dos incêndios ocorridos em Portugal continental em 1975 (a) e sua correspondente dimensão em termos de área ardida (b).*

Verifica-se que a maioria dos incêndios (e também os maiores) ocorreram mais para o norte e centro-norte do país.

Os exemplos apresentados nestes curso referentes aos processos pontuais são exclusivamente analisados recorrendo ao pacote do *software* `R`, `spatstat` (Baddeley e Turner, 2005, 2006). Este pacote permite criar, manipular, fazer gráficos de padrões espaciais e levar a cabo análises exploratórias dos dados, ajusta modelos de forma paramétrica (com base na verosimilhança) e ainda permite a simulação de modelos de processos pontuais.

Ainda no `R` há outros pacotes para a análise de processos pontuais - `spancs` (Rowlingson e Diggle, 1993; Bivand, 2001), `spatial` (Ripley, 2001), `ptproc` (Peng, 2003) e `SSlib` (Harte, 2003) - mas a nossa escolha incidiu no `spatstat` porque conjuga nele a possibilidade de levar a cabo tanto análises mais empíricas como modelações paraméticas complexas, não demasiadamente específicas da utilização do conjunto de dados que as sugerem.

4.2 Processos pontuais espaciais

4.2.1 Nomenclatura e conceitos básicos

Esta secção introduz, de forma abreviada e sem grandes formalismos, os processos pontuais em $\mathbb{R}^2$, que se apresentam como modelos matemáticos para descrever padrões aleatórios de pontos.

Os dados aqui tratados provêm de conjuntos não ordenados de pontos de $\mathbb{R}^2$, $\mathbf{x} = \{x_1, \ldots, x_n\}$, $n = 1, 2 \ldots$, que designaremos por **configurações** (ou $\mathbf{x} = \emptyset$, para a configuração vazia). Estaremos apenas interessados na família $\mathbb{N}$ das **configurações** ditas **localmente finitas**, que são tais que qualquer subconjunto limitado de $\mathbb{R}^2$ contém apenas um número finito de pontos.

No caso mais simples, um **processo pontual espacial** X definido em $\mathcal{D}$ é um subconjunto aleatório contável de um espaço $\mathcal{D} \subseteq \mathbb{R}^2$

e uma sua realização é um **padrão pontual espacial x** de pontos (configuração). Por outras palavras, um processo pontual espacial está relacionado com um mecanismo estocástico que gera um conjunto contável de ocorrências x_i de um acontecimento de interesse no plano, designados doravante por acontecimentos ou pontos.

Definição 4.1. *Define-se* ***processo pontual espacial*** X *em* $\mathcal{D} \subset \mathbb{R}^2$ *como um subconjunto aleatório localmente finito de* $\mathbb{R}^2$*. Ou seja* X *é uma aplicação do espaço de probabilidades* $(\Omega, \mathcal{S}, \mathcal{P})$ *no espaço mensurável* $(\mathbb{N}, \mathcal{N})$*, com*

$$\mathbb{N} = \{\boldsymbol{x} \subseteq \mathcal{D} : n(\boldsymbol{x}_A) < \infty,\ A \subseteq \mathcal{D},\ \textit{de Borel, limitado}\},$$

onde $n(\boldsymbol{x}_A) = N(A)$ *é o número de pontos da configuração* $\boldsymbol{x}$ *em* A *e em que se define* $\mathcal{N}$ *como a menor* σ*-álgebra de subconjuntos de* $\mathbb{N}$ *tal que para todo o conjunto* A *atrás referido, a aplicação de* $\mathcal{N}$ *em* $\mathbb{N}_0$ *que a cada* $\boldsymbol{x}$ *faz corresponder* $N_{\boldsymbol{x}}(A)$ *é mensurável.*

Como consequência da definição anterior, $N(A)$ é uma variável aleatória tomando valores finitos sempre que $A \subset \mathbb{R}^2$ for uma região limitada.

Na prática, uma representação de padrões pontuais ou mapa não contém múltiplos pontos exactamente numa mesma localização. Assim iremos aqui apenas considerar **processos pontuais simples**, para os quais $x_i \neq x_j$, se $i \neq j$.

A medida de probabilidade induzida em $\mathcal{N}$ chama-se a **distribuição de X**. Esta é então determinada pelas probabilidades:

$$P(X \in U),\ \text{para } U \in \mathcal{N}.$$

São igualmente importantes as **distribuições de dimensão finita** de um processo pontual X, que se definem como as distribuições

conjuntas do vector $\big(N(A_1), \ldots, N(A_m)\big)$ para todas as sequências de borelianos limitados $(A_1, \ldots, A_m)$ de comprimento finito $m \geq 1$:

$$P\big(N(A_1) = n_1, \ldots, P(A_m) = n_m\big),$$

com $n_i,\ i = 1, \ldots, m$, inteiros não negativos.

A distribuição de um processo pontual X em $(\mathbb{N}, \mathcal{N})$ é **completamente especificada** pelas suas distribuições de dimensão finita.

Se A for uma região com área $|\mathrm{A}|$, dizemos então que os processos que aqui consideramos, sendo processos pontuais simples, têm de obedecer à condição:

$$\lim_{|A| \to 0} \frac{P(N(A) > 1)}{|A|} = 0,$$

o que lhes confere a designação de **processos ordenados**. Este termo refere-se ao facto de se poderem ordenar os acontecimentos de forma não ambígua com respeito à sua distância a uma qualquer origem escolhida.

A distribuição de um processo pontual pode ainda ser determinada com recurso às **probabilidades nulas** de borelianos limitados $A \subseteq \mathcal{D}$:

$$v_A = P(N(A) = 0).$$

Como assumimos que o processo pontual é simples então a distribuição de X é determinada pelo conjunto de valores de v_A, percorrendo A todos os conjuntos compactos.

Trataremos usualmente de processos pontuais espaciais estacionários e isotrópicos:

Definição 4.2. *Um processo pontual espacial X diz-se* ***estacionário*** *se a sua distribuição for invariante a translações em $\mathbb{R}^2$. X diz-se ainda* ***isotrópico*** *se a sua distribuição for invariante a rotações em $\mathbb{R}^2$.*

A estacionaridade e a isotropia juntas produzem invariância ao movimento. Informalmente, a estacionaridade implica que se podem estimar as propriedades do processo de uma única realização em $\mathcal{D}$, explorando o facto dessas propriedades serem as mesmas em sub-regiões diferentes de $\mathcal{D}$. A isotropia significa que não há efeitos direccionais.

Os conceitos acima são válidos em processos observados em ambientes homogéneos e eram vulgarmente pressupostos nas primeiras aplicações dos processos pontuais espaciais, visto que se estudavam regiões consideravelmente pequenas. Abandonamos estes pressupostos sempre que exista informação sobre covariáveis espaciais, que se pensam afectar a intensidade local dos acontecimentos.

Em algumas aplicações podemos contornar o pressuposto da estacionaridade através do uso de inferência delineada para tal (por exemplo, através do uso de várias realizações do mesmo processo).

Para um processo pontual estacionário e isotrópico tem muito interesse considerar as seguintes funções distribuições associadas, respectivamente, à distância entre um ponto qualquer e um acontecimento e à distância entre um acontecimento e outro.

Define-se **função distribuição de contacto esférica** ou **função distribuição de espaço vazio**, e denota-se por F(t), a probabilidade da distância de um ponto arbitrário o (fixo) ao acontecimento mais próximo ser menor ou igual a t:

$$F(t) = 1 - P(N(b_{o,t}) = 0),$$

em que $b_{o,t}$ representa um circulo de raio t centrado em o.

Esta função pode ser interpretada como a função distribuição da distância aleatória de um ponto arbitrário (origem de $\mathbb{R}^2$) ao acontecimento mais próximo de X, ou ainda como a probabilidade de um disco de raio t aleatoriamente localizado conter pelo menos um

acontecimento.

Define-se **função distribuição do vizinho mais próximo**, e denota-se por $G(t)$, a função distribuição da distância de um acontecimento arbitrário ao acontecimento que lhe está mais próximo.

Note-se que a aplicação de $F(t)$ ou $G(t)$ na caracterização de processos pontuais é bem distinta, sendo particularmente importante para padrões com agregações - enquanto que $G(t)$ descreve os aspectos distribucionais das agregações, F(t) caracteriza em particular os espaços vazios entre elas. Esta observação motiva a importância da **função** J:

$$J(t) = \frac{1 - G(t)}{1 - F(t)}, \quad t \geq 0.$$

A maioria dos processos pontuais encontrados na prática são observados numa região limitada (quer seja esta ditada pela aplicação em causa ou porque se registam os dados numa janela de observação limitada W, menor que o espaço em estudo $\mathcal{D}$ e nele incluída, escolhida por conveniência). O mapa resultante contém então um número finito de pontos.

Assim, apesar de os modelos serem frequentemente definidos como processos em todo o plano, na prática apenas os aplicamos a dados de regiões planares limitadas, sendo tal suficiente para garantir que os nossos objectivos de estacionaridade e isotropia valem como uma aproximação razoável dentro da região em questão - as regiões de estudo são frequentemente seleccionadas com este pressuposto em mente.

No entanto, a diferença entre assumirmos que os dados $\boldsymbol{x}$ são uma realização de um processo pontual finito X definido apenas dentro de uma janela W (caso limitado) ou que são $\boldsymbol{x} \cap W$, uma realização de um processo pontual X, que se estende por todo o $\mathbb{R}^2$, observada parcialmente numa janela de observação W (caso ilimitado), tem im-

portância a nível inferencial. Neste último caso temos de lidar com o seguinte problema:

Definição 4.3. *Referem-se como* ***efeitos de fronteira (edge effects)*** *os problemas de dados faltantes causados por não se observar* $X \backslash W$.

4.2.2 Processos pontuais marcados

Associado a cada ponto x num processo espacial pontual X podemos ter uma variável aleatória m_x chamada uma *marca*. Esta marca frequentemente encerra alguma informação sobre o ponto ou outro processo pontual.

Definição 4.4. *Seja* X *um processo pontual em* $\mathcal{D} \subseteq \mathbb{R}^2$. *Dado um qualquer espaço* $\mathcal{M}$, *esteja uma "marca" aleatória* $m_x \in \mathcal{M}$ *associada a cada ponto* $x \in X$. *Então define-se um* processo pontual marcado *de pontos em* $\mathcal{D}$ *e espaço de marcas* $\mathcal{M}$ *como um processo pontual em* $\mathcal{D} \times \mathcal{M}$.

Frequentemente as marcas consistem ou num conjunto discreto de etiquetas ou a parte positiva da recta real. No caso discreto, os acontecimentos x são classificados nos diferentes tipos possíveis, indicando a marca m associada o correspondente tipo.

No caso contínuo, a marca é usualmente uma medição física, tal como, por exemplo, quando x se refere à localização de uma partícula e m o seu volume ou a sua massa, ou quando x se refere à localização de uma árvore e a marca m ao seu diâmetro do tronco ou à sua altura ou ainda quando x se refere à localização de um incêndio e m a sua dimensão em termos de, por exemplo, área ardida - veja-se Figura **??** (b).

Para estes processos as translações e as rotações actuam sobre o processo pontual marcado movendo ou rodando, respectivamente, os

pontos mas não alterando as marcas.

Um processo pontual marcado é dito estacionário (respectivamente, isotrópico) se o processo marcado transladado (respectivamente, rodado) tiver a mesma distribuição do processo marcado original.

4.2.3 Propriedades de segunda ordem

Os métodos estatísticos para padrões espaciais pontuais envolvem frequentemente a comparação entre descrições sumárias empíricas dos dados e as correspondentes descrições sumárias teóricas de um modelo de processos pontuais. *É contudo importante que as descrições sumárias teóricas sejam derivadas do modelo subjacente em vez de serem elas próprias avançadas como modelos de seu pleno direito.*

Consideramos seguidamente várias descrições sumárias teóricas de processos pontuais e as correspondentes descrições empíricas dos padrões espaciais observados. Mais tarde apresentaremos os processos pontuais que vêm a ser propostos como modelos para os dados.

Como definir para os processos pontuais o análogo às estruturas de médias e covariâncias para processos de valores reais?

Denote dx uma pequena região contendo o ponto x.

Definição 4.5. *As propriedades de primeira ordem de um processo pontual espacial são descritas através de uma* ***função intensidade (de primeira ordem)****:*

$$\lambda(x) = \lim_{|dx| \to 0} \left\{ \frac{\mathrm{E}[N(dx)]}{|dx|} \right\}.$$

Para um processo estacionário $\lambda(x) = \lambda = \dfrac{\mathrm{E}[N(A)]}{|A|}$, constante para todo o A, região limitada e não vazia de $\mathcal{D} \subseteq \mathbb{R}^2$. Temos assim, neste caso, a interpretação simples da intensidade λ como o número esperado de acontecimentos por unidade de área.

Definição 4.6. *As propriedades de segunda ordem de um processo pontual espacial são descritas através de uma* ***função intensidade de segunda ordem****:*

$$\lambda_2(x,y) = \lim_{|dx|,|dy|\to 0} \left\{ \frac{\mathrm{E}[N(dx)N(dy)]}{|dx||dy|} \right\}.$$

Para um processo estacionário, $\lambda_2(x,y) \equiv \lambda_2(x-y)$; Para um processo estacionário e isotrópico, tem-se ainda que $\lambda_2(x-y) = \lambda_2(t)$, com $t = \| x-y \|$, sendo $\| \cdot \|$ a distância euclideana.

Definição 4.7. *Definimos* ***densidade de covariância*** *de um processo pontual espacial como:*

$$\vartheta(x,y) = \lambda_2(x,y) - \lambda(x)\lambda(y).$$

Para um processo estacionário e isotrópico, $\vartheta(t) = \lambda_2(t) - \lambda^2$, com $t = \| x-y \|$.

Definição 4.8. *Define-se a intensidade num ponto x condicional à existência de um acontecimento no ponto y, designando-se tal por* ***intensidade condicional****, por:*

$$\lambda_c(x|y) = \lim_{|dx|\to 0} \left\{ \frac{\mathrm{E}[N(dx)|acontecimento\ em\ y]}{|dx|} \right\} = \frac{\lambda_2(x,y)}{\lambda(y)}.$$

Para um processo pontual espacial estacionário, isotrópico e ordenado, temos que:

$$\lambda_c(x|y) = \frac{\lambda_2(\|x-y\|)}{\lambda} = \frac{\lambda_2(t)}{\lambda}. \tag{4.1}$$

Este resultado, bem como a segunda igualdade da definição anterior, é consequência da seguinte proposição:

Proposição 4.1. *Para um processo pontual espacial ordenado temos que:*

1. $\mathrm{E}[N(dx)] \approx P(N(dx) = 1) \ \left(\Leftrightarrow \lim_{|dx|\to 0} \frac{\mathrm{E}[N(dx)]}{P(N(dx)=1)} = 1\right)$;

2. $\mathrm{E}[N(dx)N(dy)] \approx P(N(dx) = N(dy) = 1)$.

Definição 4.9. *Dizemos que temos um processo pontual espacial estacionário (de segunda ordem) reponderado se:*

$$\frac{\lambda_2(x,y)}{\lambda(x)\lambda(y)} = \rho(t), \quad \textit{depende só de } t = \| x - y \|.$$

Note-se a estacionaridade reponderada é o análogo nos processos pontuais ao pressuposto comum na análise de processos espaciais de valores reais de que a média pode variar espacialmente enquanto que a variação em torno da média local é estacionária.

A partir daqui, salvo explícita menção em contrário, assumimos sempre estacionaridade e isotropia (e ordenação). Nestas condições já vimos que a intensidade de primeira ordem tem uma interpretação fácil, não acontecendo o mesmo com a de segunda ordem. Assim surge a necessidade do seguinte:

Definição 4.10. *Uma caracterização alternativa das propriedades de segunda ordem de um processo estacionário e isotrópico é dada pela função de segundo momento reduzido, $K(t)$, dada por:*

$$K(t) = 2\pi\lambda^{-2} \int_0^t \lambda_2(u) u \, du.$$

A relação entre $K(t)$ e $\lambda_2(t)$ para um processo pontual espacial estacionário, isotrópico e ordenado pode-se ainda escrever como:

$$\lambda K(t) = 2\pi\lambda^{-1} \int_0^t \lambda_2(u) u \, du \quad \Leftrightarrow \quad \lambda_2(t) = \lambda^2 (2\pi t)^{-1} K'(t) \quad (4.2)$$

A real importância da definição anterior é que $K(t)$ pode ser expresso como o valor esperado de uma quantidade observável, que por sua vez sugere como o poderemos estimar dos dados:

Proposição 4.2. *Sendo $N_0(t)$ é o número de outros acontecimentos dentro de uma distância t de um acontecimento arbitrário, temos que*

$$K(t) = \lambda^{-1}\mathrm{E}[N_0(t)],$$

Demonstração:

Da Proposição 4.1 decorre que o número esperado de outros acontecimentos dentro de uma distância t de um acontecimento arbitrário, $\mathrm{E}[N_0(t)]$, pode ser calculado pela integração da intensidade condicional sobre um disco de centro, por exemplo, na origem, o, e raio t:

$$\begin{aligned}
\mathrm{E}[N_0(t)] &= \int_0^{2\pi}\int_0^t \lambda_c(r|o)r\,dr\,d\theta = \qquad \text{(por (4.8))}\\
&= \frac{2\pi}{\lambda}\int_0^t \lambda_2(r)r\,dr = \qquad \text{(pela definição de } K(t))\\
&= \frac{2\pi}{\lambda}\frac{\lambda^2 K(t)}{2\pi} = \lambda K(t).
\end{aligned}$$

■

Observações:

1. Observe-se que a proposição anterior serve por vezes de definição da função $K(t)$ sendo, nesse caso, a definição aqui apresentada uma sua consequência.
2. A proposição anterior sugere porque é que uma estimativa de $K(t)$ seria útil para descrever um padrão espacial observado:

- para padrões em que os acontecimentos têm grande probabilidade de estar rodeados de um espaço vazio, o número esperado de outros acontecimentos dentro de uma pequena distância será relativamente pequeno;
- para acontecimentos que tendem a aparecer aglomerados, o número esperado de outros acontecimentos dentro de uma pequena distância será relativamente grande.

3. De um ponto de vista prático a vantagem de trabalhar com $K(t)$ em vez de $\lambda_2(t)$ é que o primeiro é mais facilmente estimado de forma automática dos dados;
4. A função K é invariante sob estreitamentos aleatórios. Ou seja, se cada acontecimento do processo for retido ou não de acordo com uma série de provas Bernoulli independentes, então a função K do processo resultante estreitado é idêntica á do processo inicial.

Combinando as definições de $\vartheta(t)$ com a definição de $K(t)$ obtemos:

$$\begin{aligned} K(t) &= 2\pi\lambda^{-1}\int_0^t (\vartheta(u)+\lambda^2)u\,du = \\ &= \pi t^2 + 2\pi\lambda^{-2}\int_0^t \vartheta(u)u\,du. \end{aligned} \tag{4.3}$$

Assim a função πt^2 fornece uma marca de referência importante na interpretação de $K(t)$. Para padrões regulares veremos que $K(t) < \pi t^2$ e o contrário para padrões agregados.

Definição 4.11. *Para um processo estacionário reponderado tem-se a seguinte extensão da definição da função K:*

$$K_I(t) = 2\pi\int_0^t \rho(x)x\,dx,$$

que se reduz à primeira equação de 4.2, no caso de estacionaridade.

4.2.4 Propriedades de ordem superior e distribuições de vizinhos mais próximos

As propriedades de segunda ordem são um bom ponto de partida, mas devemos lembrar-nos que podem existir dois processos pontuais completamente distintos com exactamente as mesmas propriedades de segunda ordem. No entanto, as propriedades de ordem superior a dois para descrever os processos pontuais espaciais podem ser bastante difíceis de interpretar.

Uma alternativa possível é o uso das seguintes duas funções distribuição já anteriormente definidas, que servem como descrições sumárias adicionais dos processos pontuais espaciais:

1. $F(t)$, a função distribuição de contacto esférica, função distribuição da distância de um ponto arbitrário ao acontecimento que lhe está mais próximo;
2. $G(t)$, a função distribuição da distância de um acontecimento arbitrário (que podemos situar na origem) ao acontecimento que lhe está mais próximo.

A expressão destas funções é, na maioria das situações, analiticamente intratável, o que limita um pouco a sua utilidade enquanto ferramentas para a construção de modelos.

4.2.5 O Processo de Poisson homogéneo

Dizemos que um padrão espacial é completamente aleatório se for uma realização de um processo de Poisson homogéneo. Este é o modelo estocástico mais simples e importante na modelação de padrões espaciais pontuais, no qual assenta toda a teoria dos processos pontuais:

Definição 4.12. *Um* ***Processo de Poisson homogéneo****, designado daqui em diante simplesmente por processo de Poisson, caracteriza-se pelas seguintes propriedades:*

PP1 *Para algum $\lambda > 0$ e uma qualquer região finita planar A, $N(A)$ segue uma distribuição de Poisson com média $\lambda|A|$.*

PP2 *Para quaisquer duas regiões disjuntas A e B, as variáveis aleatórias $N(A)$ e $N(B)$ são independentes.*

O parâmetro λ do processo de Poisson é a sua intensidade.

A aplicação do processo de Poisson homogéneo usa-se como um padrão idealizado de *aleatoriedade espacial completa* que, ainda que não seja atingível na prática, tem um papel central como modelo nulo e como ponto de partida para a construção de modelos realísticos para os padrões pontuais.

Proposição 4.3. *O processo de Poisson é estacionário e isotrópico.*

Proposição 4.4. *Para um processo de Poisson tem-se que:*

$$\lambda_2(t) = \lambda^2, \quad t > 0$$

e, consequentemente, também que:

$$K(t) = \pi\ t^2, \quad t > 0$$

Demonstração:

$$\begin{aligned}
\lambda_2(x,y) &= \lim_{|dx|,|dy|\to 0} \left\{ \frac{\mathrm{E}[N(dx)N(dy)]}{|dx||dy|} \right\} = \qquad \text{(por } \mathbf{PP2}\text{)} \\
&= \lim_{|dx|,|dy|\to 0} \left\{ \frac{\mathrm{E}[N(dx)] \times \mathrm{E}[N(dy)]}{|dx||dy|} \right\} = \\
&= \lim_{|dx|\to 0} \left\{ \frac{\mathrm{E}[N(dx)]}{|dx|} \right\} \times \lim_{|dy|\to 0} \left\{ \frac{\mathrm{E}[N(dy)]}{|dy|} \right\} = \\
&= \lambda(x) \times \lambda(y) = \lambda^2, \qquad \text{por estacionaridade.}
\end{aligned}$$

A expressão de $K(t)$ segue então da primeira equação de (4.2).

■

Este resultado sobre $K(t)$ era o esperado, dada a sua interpretação (Proposição 4.2). Os acontecimentos num processo de Poisson estão localizados independentemente uns dos outros. Assim, a existência de um acontecimento em x não influencia o número de acontecimentos extra dentro de uma distância t de x. Sendo λ o número esperado de acontecimentos por unidade de área, o número esperado de outros acontecimentos dentro de uma distância t de x, $\mathrm{E}[N_0(t)]$, é $\lambda\pi t^2$.

Sendo o processo de Poisson o padrão de aleatoriedade espacial completa, o ponto de referência para a comparação de $K(t)$, πt^2, atrás descrito, é agora evidente.

Proposição 4.5. *Para um processo de Poisson as funções distribuição do vizinho mais próximo $G(t)$ e $F(t)$ são iguais dadas por:*

$$F(t) = G(t) = 1 - \exp(-\pi\lambda t^2), \quad t > 0. \tag{4.4}$$

Demonstração:

Este resultado decorre de **PP1** e do facto de a existência de um acontecimento num determinado ponto, digamos x_0, não ter efeito na distribuição do número restante de acontecimentos num disco com o centro em x_0:

$$\begin{aligned} F(t) = G(t) &= P(N(\pi t^2) > 0) = \\ &= 1 - P(N(\pi t^2) = 0) = 1 - \exp(-\lambda\pi t^2). \end{aligned}$$

■

Proposição 4.6. *Consideremos a restrição de um processo de Poisson X a um conjunto compacto $A \subseteq \mathbb{R}^2$, sob a condição de que $N(A) = n$. O processo pontual formado por estes n pontos em A tem a mesma distribuição do que n pontos independentes e uniformemente distribuídos em A.*

É a partir deste resultado que a simulação de um processo de Poisson se faz. Se estivermos interessados em fazê-lo numa janela W teremos primeiro de simular uma variável aleatória de Poisson para determinar o número n de pontos que aí se vão encontrar e depois as suas posições podem ser obtidas pela geração independente de n pontos de uma distribuição Uniforme em W.

Na Figura 4.8 apresentam-se três realizações simuladas do processo de Poisson numa janela quadrada de lado unitário (100 pontos). Repare-se que que este processo pode levar à formação de padrões mais ou menos agregados.

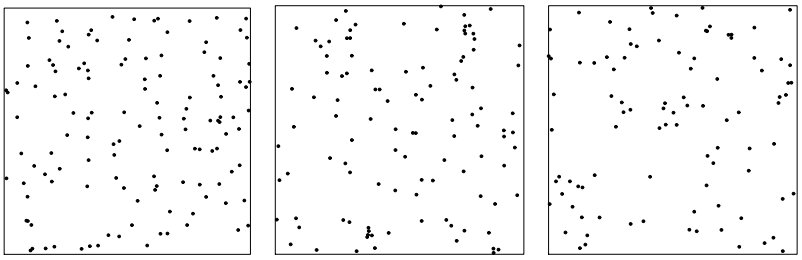


Figura 4.4 *Três realizações simuladas do processo de Poisson (100 pontos).*

Estes processos de Poisson são fundamentais enquanto processos de referência, para ferramentas exploratórias e de diagnóstico e também

quando modelos de processos pontuais espaciais mais avançados são construídos.

4.2.6 O Processo de Poisson não homogéneo

Quando a intensidade constante λ de um processo de Poisson é substituída por uma intensidade espacialmente variável, $\lambda(x)$, obtém-se uma classe de processos pontuais não estacionários, processos de Poisson não homogéneos.

Definição 4.13. *Um processo pontual diz-se um* ***Processo de Poisson não homogéneo*** *se:*

NH1 *O número de acontecimentos numa dada região planar do espaço A, $N(A)$, segue uma distribuição de Poisson com média dada por $\int_A \lambda(x)\,dx$, para uma qualquer função não-negativa $\lambda(x)$.*

NH2 *Dado que $N(A) = n$, os n acontecimentos formam uma amostra aleatória da distribuição em A com função densidade probabilidade proporcional a $\lambda(x)$.*

NH3 *Para duas quaisquer regiões disjuntas A e B, as variáveis aleatórias $N(A)$ e $N(B)$ são independentes.*

Estes processos de Poisson não homogéneos providenciam uma moldura possível para a inclusão de uma tendência espacial (modificação na função intensidade sobre a região de observação), $\lambda(x) = \lambda(x;\theta)$, ou a introdução de covariáveis $z_1(x), z_2(x), \ldots, z_p(x)$ na análise dos padrões espaciais, via a função intensidade:

$$\lambda(x) = \lambda(z_1(x), z_2(x), \ldots, z_p(x)).$$

Proposição 4.7. *Um processo de Poisson não homogéneo satisfaz o seguinte:*

1. *É estacionário de segunda ordem repesado, desde que* $\lambda(x)$ *esteja convenientemente limitado a não tomar o valor zero.*
2. *A sua intensidade de segunda ordem repesada é dada por* $\rho(t) = 1$.
3. *A sua função* K *repesada é dada por* $K_I(t) = \pi t^2$.

Um processo de Poisson não homogéneo com função intensidade $\lambda(x)$ produzirá aglomerados aparentes de acontecimentos em regiões com intensidade relativamente alta - os aglomerados obtidos desta forma prendem-se com a heterogeneidade ambiental. Veja-se o exemplo da Figura 4.5 em que se representa o padrão espacial obtido em Portugal continental por simulação de um processo de Poisson não homogéneo em que a função intensidade é crescente com a latitude, $\lambda(x, y) \propto y$:

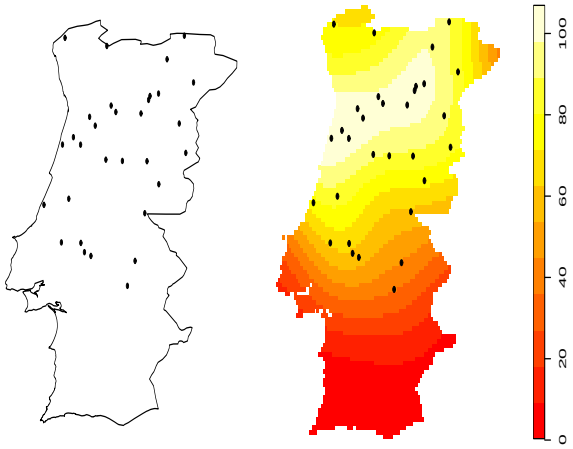


Figura 4.5 *Realização simulada de um processo não homogéneo com intensidade* $\lambda(x, y) \propto y$, *em Portugal continental.*

4.2.7 Aplicações

Qualquer estudo de padrões espaciais deve começar com uma análise exploratória dos dados. Para este tipo de dados é aconselhado iniciar esta análise com uma avaliação da homogeneidade espacial do fenómeno em estudo, usando ferramentas como estimativas alisadas kernel da intensidade do fenómeno (Diggle, 1985) - disponíveis no `spatstat` como função `ksmooth.ppp` - ou como métodos LISA (Local Indicators of Spatial Association), Anselin (1995).

Na Figura 4.6 mostram-se os gráficos das estimativas kernel da intensidade dos padrões pontuais observados nos casos das mudas de pinheiro negro Japonês e dos rebentos de Pau-Brasil, obtidos no `spatstat`, bem como os próprios pontos observados. Verifica-se não parecer haver tendências na intensidade estimada em qualquer dos casos.

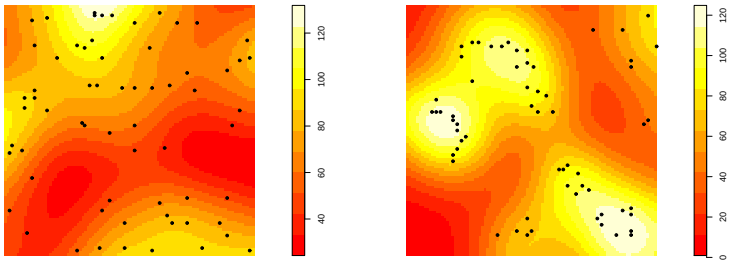


Figura 4.6 *Estimativas kernel da intensidade dos padrões espaciais observados no caso das mudas de pinheiro negro Japonês (esquerda) e dos rebentos de Pau-Brasil (direita).*

Na Figura 4.7 apresentamos o mesmo tipo de gráfico para o exemplo

das localizações dos incêndios registados em Portugal continental em 1975. Equacionamos se haverá falta de homogeneidade neste caso, claramente a intensidade é muito superior no centro-norte e norte do país.

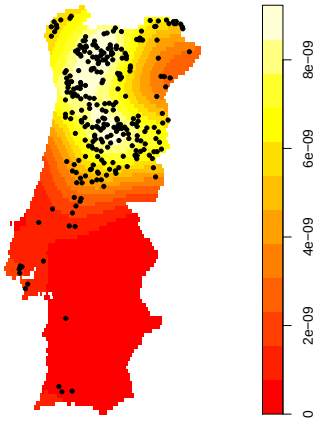


Figura 4.7 *Estimativas kernel da intensidade dos padrões espaciais observados no caso dos incêndios em Portugal continental no ano de 1975.*

No caso de se considerarem os dados espacialmente homogéneos, o próximo passo será a análise exploratória usando estatísticas resumo padrão tal como a função K, por exemplo, que devem ser estimadas dos dados.

Se os dados forem considerados espacialmente não-homogéneos, então há um grande limite ao que se pode fazer por agora em termos de análise exploratória. Uma excepção a esta observação é a função K não homogénea, implementada no `spatstat` como `Kinhom`, que ainda contém algumas outras funções que generalizam as técnicas

exploratórias aos processos não homogéneos.

4.2.8 Estimação das propriedades de segunda ordem

Seguidamente apresentamos a estimação não-paramétrica de algumas das medidas resumo mais importantes dos processos pontuais antes descritas, ferramentas úteis para se fazer uma análise exploratória dos dados. Note-se, contudo, que estas características ainda têm um papel muito importante no caso paramétrico, em que se usam modelos de processos pontuais.

Nos processos pontuais frequentemente a análise estatística assenta em apenas uma amostra do processo e tal através de uma janela limitada W. Assim tem de se assumir que os padrões observados são amostras de um processo pontual estacionário e ergódico, para que a análise estatística possa ser levada a cabo.

Na prática, ou os pressupostos acima são plausíveis para os dados ou temos de prosseguir numa base *ad hoc* de que o pressuposto é verdade, sujeitando as nossas conclusões à reserva de que apesar das características resumo poderem ser um valor possível, no caso não estacionário, não terão a mesma interpretação.

A ênfase da estimação será na função K, já que se tivermos uma sua estimativa podemos sempre usar a segunda equação de (4.2) para determinar a correspondente estimativa de $\lambda_2(t)$.

Sendo $\hat{K}(t)$ a estimativa de $K(t)$, escolhemos uma largura de banda $h > 0$, fazemos a aproximação

$$\hat{K}'(t) \approx \frac{\hat{K}(t+h) - \hat{K}(t)}{h}$$

e deduzimos a estimativa de $\lambda_2(t)$:

$$\hat{\lambda}_2(t) = \hat{\lambda}^2 (2\pi t)^{-1} \hat{K}'(t),$$

produzindo uma estimativa para $\lambda_2(t)$ tipo histograma em intervalos h de t.

Estimação de $K(t)$

Vimos atrás que a função $K(t)$ pode ser dada por $\lambda K(t) = \mathrm{E}[N_0(t)]$.

Representando a intensidade λ o número médio de acontecimentos por unidade de área, um seu estimador óbvio é o número de acontecimentos observados por unidade de área:

$$\hat{\lambda} = \frac{n}{|W|}.$$

Sendo $\mathrm{E}(t) = \mathrm{E}[N_0(t)]$ o número esperado de outros acontecimentos dentro de uma distância t de um acontecimento arbitrário e $d_{ij} = \| x_i - x_j \|$ a distância entre os acontecimentos x_i e x_j, definimos para $\mathrm{E}(t)$ o seguinte estimador:

$$\tilde{\mathrm{E}}(t) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i}^{n} I(d_{ij} \leq t)$$

em que $I(\cdot)$ denota a função indicatriz. Este estimador obtém-se simplesmente substituindo um valor esperado pela correspondente média amostral.

Este estimador é enviesado negativamente por causa dos efeitos fronteira - para um qualquer acontecimento que esteja a uma distância da fronteira de W inferior a t, o número observado de outros acontecimentos que dele distem menos de t não contabiliza necessariamente aqueles acontecimentos que estão nessas circunstâncias mas que caem fora de W.

Há várias formas de corrigir este problema, baseadas na inclusão de pesos. Por exemplo (Ripley, 1976):

- Defina-se w_{ij} como a proporção da circunferência centrada em x_i e raio dado pela distância d_{ij} que cai dentro de W.

  Para um processo estacionário e isotrópico w_{ij} é a *probabilidade condicional de um acontecimento ser observado, dado que se encontra a uma distância d_{ij} do acontecimento x_i.*

  Note-se que w_{ij} poderá ser diferente de w_{ji}, desde que qualquer uma das correspondentes circunferências não esteja totalmente dentro de W.

- Temos então o seguinte estimador centrado de $\mathrm{E}(t)$:

$$\hat{\mathrm{E}}(t) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{I(d_{ij}\leq t)}{w_{ij}}.$$

Desta forma obtemos o seguinte estimador para $K(t)$, tornando claro que função K está intimamente relacionada com a distribuição da distância inter-acontecimentos:

$$\begin{aligned}\hat{K}(t) &= \frac{1}{n^2}|W|\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{I(d_{ij}\leq t)}{w_{ij}} \approx \\ &\approx \frac{1}{n(n-1)}|W|\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{I(d_{ij}\leq t)}{w_{ij}}, \qquad (4.5)\end{aligned}$$

sendo a aproximação numericamente indiferente para valores grandes de n, mas tecnicamente conveniente.

Este estimador é aproximadamente centrado para valores de t suficientemente pequenos, sendo esta restrição sobre t necessária porque os pesos $\{w_{ij}^{-1}\}$ podem vir ilimitados à medida que o t aumenta - o que não constitui problema na prática.

Verifica-se que os maiores pesos $\{w_{ij}^{-1}\}$ tenderão a estar associados a pares de pontos mais distanciados, pelo que $\mathrm{Var}(\hat{K}(t))$ tipicamente tenderá a aumentar com t.

A determinação dos pesos pode ser explicitamente obtida para formas simples da região W, como por exemplo quando W é rectangular ou circular (e.g. Diggle, 2003, página 51).

A distribuição por amostragem de $\hat{K}(t)$ é intratável analiticamente, excepto no caso de um processo de Poisson homogéneo. Dado um qualquer modelo e uma qualquer região W, a distribuição por amostragem de $\hat{K}(t)$ pode ser estimada por simulação directa.

Contudo a expressão teórica da variância de $\hat{K}(t)$ num processo de Poisson homogéneo pode ser útil para uma inspecção inicial de um gráfico de $\hat{K}(t)$. Apresentam-se de seguida algumas alternativas para tal, onde se assume que o número de acontecimentos n em W é fixo:

- Aproximação assimptótica de Ripley (1988) (precisa para valores relativamente pequenos de t):

$$v_R(t) = 2\left(\frac{|W|}{n-1}\right)^2\left(\frac{\pi t^2}{|W|} + \frac{0.96Pt^3}{|W|^2} + \frac{0.13nPt^5}{|W|^3}\right),$$

  sendo P o perímetro de W.

- Forma exacta de Lotwick e Silverman (1982), assumindo que W é rectangular:

$$v_{LS}(t) = \left(\frac{n-1}{n^3}\right)|W|^2\big(2b(t) - a_1(t) + (n-2)a_2(t)\big),$$

  onde:

$$\begin{aligned} b(t) &= \pi t^2|W|^{-1}(1-\pi t^2/|W|) + |W|^{-2}(1.0716Pt^3 + 2.2375t^4),\\ a_1(t) &= |W|^{-2}(0.21Pt^3 + 1.3t^4) \quad \text{e}\\ a_2(t) &= |W|^{-3}(0.24Pt^5 + 2.62t^6), \end{aligned}$$

  fórmulas válidas para t menor ou igual que o menor lado do rectângulo W.

- Aproximação de Chatwynd e Diggle (1998) para qualquer forma W:

$$v_{CD}(t) = \left(\frac{2|W|^2}{n^{(2)}}\right)((3-2n)m_4(t) + 2(n-2)m_3(t) + m_2(t)),$$

onde:

$$n^{(k)} = n(n-1)\ldots(n-k+1);$$
$$m_2(t) = \frac{X_n}{n^{(2)}}; \quad m_3(t) = \frac{Z_n - X_n}{n^{(3)}}; \quad m_4(t) = \frac{W_n^2 - 4Z_n + 2X_n}{n^{(4)}};$$
$$W_n = \sum_{i=1}^{n}\sum_{j\neq i}^{n}\phi_{ij}; \quad X_n = \sum_{i=1}^{n}\sum_{j\neq i}^{n}\phi_{ij}^2; \quad Z_n = \sum_{i=1}^{n}\left(\sum_{j\neq i}^{n}\phi_{ij}\right)^2;$$
$$\phi_{ij} = 0.5(w_{ij} + w_{ji})I(\| x_i - x_j \| \leq t).$$

Assim, para uma análise exploratória da estrutura de segunda ordem de padrões próximos da aleatoriedade espacial completa, é útil colocar em gráfico as estimativas $\hat{K}(t)$, eventualmente com limites de erros baseados nas expressões acima - por exemplo v_{LS} para regiões rectangulares e v_{CD} para regiões irregulares.

No entanto para padrões claramente diferentes da aleatoriedade espacial completa estas aproximações baseadas na Poisson não são de confiança. Quando se formula um modelo paramétrico, os erros padrão podem ser estimados por simulação do referido modelo. Na ausência deste pode recorrer-se ao seguinte estratagema:

Subdividir W em m sub-áreas iguais, estimar $K(t)$ separadamente em cada sub-área i (designe-se por k_i o correspondente estimador aí) e usar a variância empírica sobre estes. Assim obtém-se o estimador global:

$$\tilde{K}(t) = \frac{1}{m}\sum_{i=1}^{m} k_i,$$

com variância aproximada dada por:

$$\mathrm{Var}(\tilde{K}(t)) = \frac{1}{m(m-1)}\sum_{i=1}^{m}(k_i - \tilde{K}(t))^2.$$

Note-se que este estimador de $K(t)$ espera-se menos eficiente que $\hat{K}(t)$, já que não usa a informação de pares de acontecimentos que se encontrem em sub-áreas distintas, o que pode ser problemático para valores pequenos de n. Além disso na estimação da sua variância assumem-se independentes os estimadores k_i entre sub-áreas, o que só será verdade no caso de um processo de Poisson homogéneo.

Assim sugere-se que o uso deste estimador seja restrito ao caso de pequenos valores de t ou quando a divisão artificial de W seja substituída por replicação genuína.

Proposição 4.8. *No caso de o processo não ser homogéneo, ser estacionário repesado e, pouco realisticamente, a sua intensidade de primeira ordem $\lambda(x)$ ser conhecida, pode-se estimar $K_I(t)$ fazendo em (4.5) uma modificação que consiste num re-escalonamento das distâncias inter-acontecimentos pelo produto das intensidades de primeira ordem nas correspondentes duas localizações:*

$$\hat{K}_I(t;\lambda) = \frac{1}{|W|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{I(d_{ij}\leq t)}{w_{ij}\lambda(x_i)\lambda(x_j)}$$

Quando se tenta usar um estimador da intensidade de primeira ordem $\hat{\lambda}(x)$ em vez da própria intensidade $\lambda(x)$ acaba-se por ter bastantes problemas na prática - por causa da dificuldade em distinguir empiricamente entre a não constância de $\lambda(x)$ e a dependência entre acontecimentos do processo. É portanto muito difícil, na prática, estimar não parametricamente em simultâneo as propriedades de primeira e de segunda ordens de um processo estacionário repesado.

4.2.9 Aplicações

Relativamente ao exemplo das mudas de pinheiros negros Japoneses, apresentam-se de seguida as estimativas da função K. Os gráficos foram construídos recorrendo ao pacote do `R` `spatstat` - função `Kest`.

Na Figura 4.8 temos, à esquerda, a função K estimada (linha sólida) e a função K sob a validade da hipótese de aleatoriedade espacial completa (a tracejado) e à direita representa-se a função K estimada menos o correspondente valor teórico desta função (linha sólida) e também limites dados, sob a validade da hipótese de aleatoriedade espacial completa, por mais e menos 2 desvios padrão calculados com base nas fórmulas de Lotwick-Silverman. Repare-se que, neste último gráfico, o valor da estimativa se encontra dentro dos limites teóricos, sugerindo compatibilidade dos dados com a hipótese de aleatoriedade espacial completa.

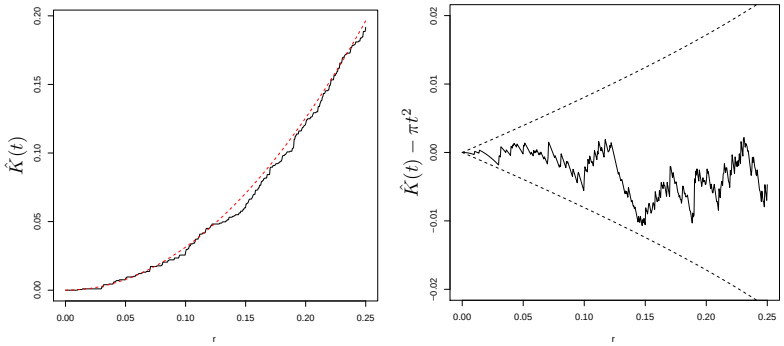


Figura 4.8 $\hat{K}(t)$ *(linha a cheio) e a correspondente função teórica sob a hipótese de aleatoriedade espacial completa* ($K(t) = \pi t^2$) *(esquerda) e* $\left(\hat{K}(t) - \pi t^2\right)$ *e correspondentes 'limites de confiança' (direita), para os dados das mudas de pinheiro negro Japonês.*

Apresentam-se agora as estimativas da função K para o exemplo dos dados dos rebentos do Pau-Brasil - Figura 4.9. Repare-se que aqui claramente os dados são incompatíveis com a hipótese de aleatoriedade espacial completa - certamente por causa do padrão agregado que já tivemos oportunidade de observar anteriormente.

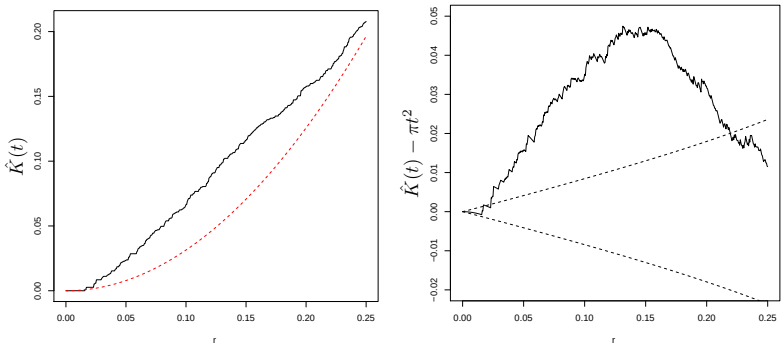


Figura 4.9 $\hat{K}(t)$ *(linha a cheio) e a correspondente função teórica sob a hipótese de aleatoriedade completa* ($K(t) = \pi t^2$) *(esquerda) e* $\left(\hat{K}(t) - \pi t^2\right)$ *e correspondentes 'limites de confiança' (direita), para os dados dos rebentos de Pau-Brasil.*

Na Figura 4.10 apresentamos as estimativas da função K para o padrão espacial simulado de um processo de Poisson não homogéneo representado na Figura 4.5. Do lado esquerdo encontra-se a estimativa assumindo a homogeneidade do processo subjacente e do lado direito a falta da mesma.

Verifica-se que a estimativa da função K, considerando incorrectamente que existe homogeneidade nos dados, está mais distante do verdadeiro valor teórico desta função sob a hipótese de aleatoriedade espacial completa. Mais, esta estimativa sobre-estima esta função, com as consequências óbvias que decorrem da proposição (4.2).

Na Figura 4.11 dispõem-se as estimativas da função K para os dados das localizações dos fogos registados em 1975 em Portugal continental. Do lado esquerdo fazemos a estimação ignorando alguma

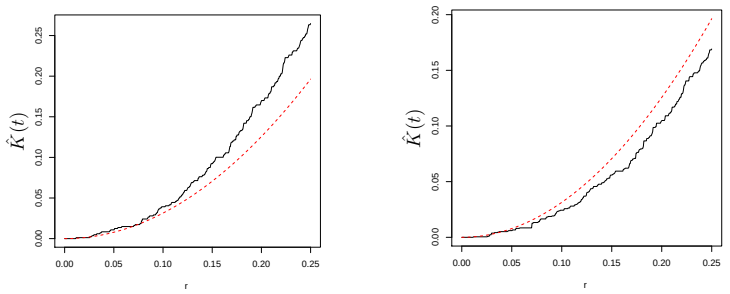


Figura 4.10 $\hat{K}(t)$ *(linha a cheio) e a correspondente função teórica sob a hipótese de aleatoriedade espacial completa* ($K(t) = \pi t^2$)*, assumindo homogeneidade (esquerda) e falta de homogeneidade (direita), para os dados simulados representados na Figura 4.5.*

falta de homogeneidade dos dados e do lado direito consideramos que os dados não são homogéneos.

4.2.10 Estimação das funções distribuição do espaço vazio e do vizinho mais próximo

Estas funções são naturalmente bem estimadas através das correspondentes funções distribuição empíricas.

Para a estimação da função distribuição do espaço vazio, $F(t)$, cubra-se W com uma grelha fina de m pontos e represente d_i a distância do i-ésimo *ponto* ao acontecimento mais próximo. Então, a correspondente função distribuição empírica é:

$$\tilde{F}(t) = \frac{\sum_{i=1}^{m} I(d_i \leq t)}{m}. \tag{4.6}$$

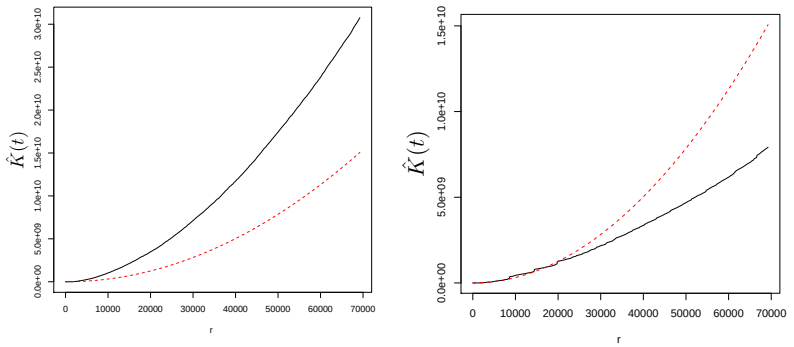


Figura 4.11 $\hat{K}(t)$ *(linha a cheio) e a correspondente função teórica sob a hipótese de aleatoriedade espacial completa* ($K(t) = \pi t^2$)*, assumindo homogeneidade (esquerda) e falta de homogeneidade (direita), para os dados da localizações dos incêndios em Portugal continental em 1975.*

Por causa dos efeitos fronteira este estimador é enviesado, podendo uma possível solução ser a de usar apenas aqueles pontos que se encontram a pelo menos uma distância t da fronteira de W.

Definindo assim e_i como a distância do $i-$ésimo ponto ao ponto da fronteira de W que lhe está mais próximo, temos o seguinte estimador centrado de $F(t)$:

$$\hat{F}(t) = \frac{\sum_{i=1}^{m} I(e_i \geq t, d_i \leq t)}{\sum_{i=1}^{m} I(e_i \geq t)}. \tag{4.7}$$

O aumento do número de pontos na grelha m não é problema hoje em dia com as facilidades computacionais à nossa disposição, o que permite encontrarmos uma curva empírica $\hat{F}$ bastante alisada. No entanto não nos devemos esquecer que a precisão estatística deste estimador continua limitada por n, o número de acontecimentos.

Para a função distribuição do vizinho mais próximo, $G(t)$, usamos o mesmo tipo de estimação, mas medindo as distâncias d_i do i-ésimo *acontecimento* ao acontecimento que lhe está mais próximo e as distâncias e_i do $i-$ésimo acontecimento ao ponto que lhe está mais próximo na fronteira de W. Assim temos:

$$\tilde{G}(t) = \frac{\sum_{i=1}^{n} I(d_i \leq t)}{n}. \tag{4.8}$$

e

$$\hat{G}(t) = \frac{\sum_{i=1}^{n} I(e_i \geq t, d_i \leq t)}{\sum_{i=1}^{n} I(e_i \geq t)}. \tag{4.9}$$

Na prática, se o nosso objectivo é apenas o de comparar estimativas calculadas a partir dos dados com estimativas calculadas da simulação de modelos hipotéticos, os efeitos fronteira podem ser ignorados e os estimadores $\tilde{F}(t)$ e $\tilde{G}(t)$ usados.

4.2.11 Aplicações

Na Figura 4.12 apresentam-se as estimativas das funções distribuição do espaço vazio e do vizinho mais próximo (linhas a cheio), para os dados das mudas de pinheiro negro Japonês. Na Figura 4.13 apresentam-se as mesmas estimativas para os dados dos rebentos de Pau-Brasil e na Figura 4.14 para os dados das localizações dos incêndios em Portugal Continental em 1975. A linha a tracejado corresponde à correspondente função teórica sob a validade da hipótese de aleatoriedade espacial completa. Estes gráficos foram produzidos com recurso às funções `Fest` e `Gest` do pacote `spatstat` do `R`.

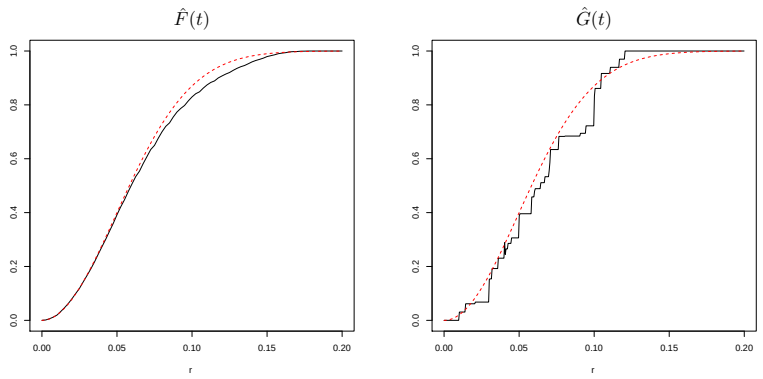


Figura 4.12 *Funções distribuição empíricas das distâncias de vizinhos (a cheio), para as mudas de pinheiro negro Japonês.*

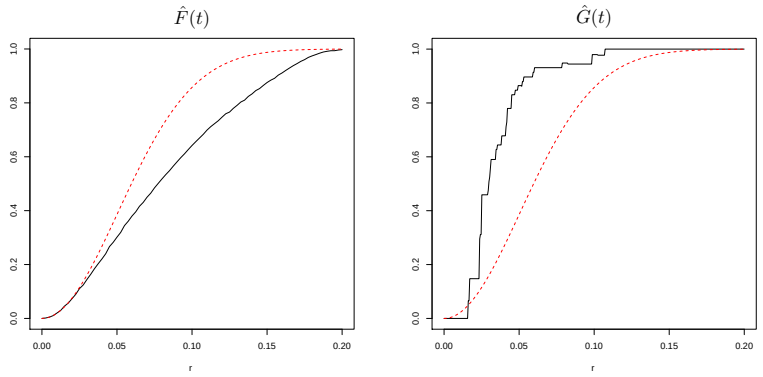


Figura 4.13 *Funções distribuição empíricas das distâncias de vizinhos (a cheio), para os rebentos de Pau-Brasil.*

4.3 Teste à hipótese de aleatoriedade espacial completa

A hipótese de aleatoriedade espacial completa é usada para dividir padrões em regulares ou agregados e os testes estatísticos a ela asso-

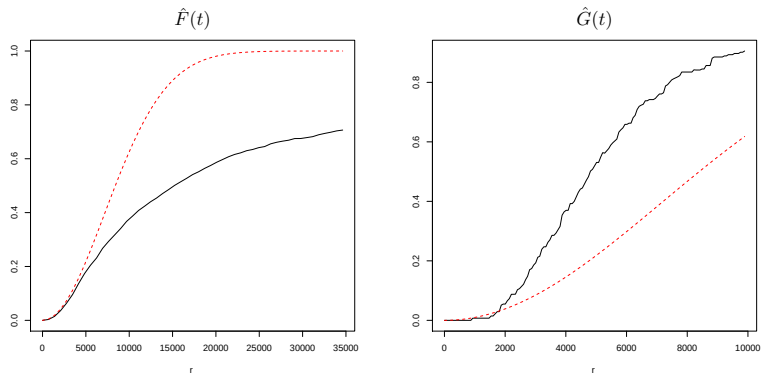


Figura 4.14 *Funções distribuição empíricas das distâncias de vizinhos (a cheio), para as localizações dos fogos em Portugal Continental em 1975.*

ciados servem de forma de exploração das configurações observadas, ajudando na formulação de modelos alternativos adequados.

Estes testes devem ser encarados como análise preliminar da modelação dos padrões espaciais pontuais. São testes usando a metodologia de simulação Monte Carlo e fazem parte das primeiras ferramentas estatísticas usadas na análise dos processos pontuais.

Teste usando distâncias ao vizinho mais próximo

Em muitas situações práticas, se existem interacções entre acontecimentos, elas existem em pequena escala, para aquelas ocorrências que se encontram próximas no espaço - por exemplo as plantas competem por nutrientes com aquelas outras plantas que a cercam. Nestes casos, as distâncias entre acontecimentos vizinhos fornecem uma forma objectiva de nos concentrarmos na interacção de pequena escala.

Podemos então comparar a distribuição destas distâncias $G(t)$, sob

a hipótese de aleatoriedade espacial completa, (4.4), com a correspondente distribuição empírica, $\hat{G}(t)$, (4.8) ou (4.9), através de gráfico de uma contra a outra - esperando obter, no caso de valer a hipótese, um gráfico aproximadamente linear. Por facilidade de exposição do que se segue denotamos nesta secção $\hat{G}(t)$ por $\hat{G}_1(t)$.

Para avaliar a significância ou outras formas de desvio da linearidade, a abordagem convencional seria a consideração da distribuição por amostragem de $\hat{G}(t)$, sob a validade da hipótese de aleatoriedade espacial completa, mas tal tarefa não se revela simples.

Assim, recorre-se a uma abordagem baseada em simulação de Monte Carlo. Calculam-se então funções distribuição empíricas $\hat{G}_i(t)$, $i = 2, 3 \ldots, s$, para cada uma de $(s-1)$ simulações independentes de n acontecimentos independentes e uniformemente distribuídos em W, e definem-se invólucros de simulação inferior e superior, respectivamente, por [1]:

$$L(t) = \min\{\hat{G}_i(t), i = 2, 3 \ldots, s\}$$
$$U(t) = \max\{\hat{G}_i(t), i = 2, 3 \ldots, s\}$$

Também podemos pôr estes invólucros em gráfico contra $G(t)$. Além disso têm a propriedade de, para cada t, e sob a validade da hipótese de aleatoriedade espacial completa,

$$P\big(\hat{G}_1(t) < L(t)\big) = P\big(\hat{G}_1(t) > U(t)\big) = \frac{1}{s}.$$

Uma forma de fazer um teste de Monte Carlo à hipótese de aleatoriedade espacial completa é usar a seguinte medida de discrepância u_i entre as distribuições empíricas (dos dados e das simulações) e a

[1] Chamamos a atenção que na construção dos invólucros por simulação Monte Carlo utiliza-se sempre a mesma metodologia - conferir com a subsecção 2.3.2.

verdadeira distribuição de $G(t)$ e verificar se a dos dados é superior à das simulações:

$$u_i = \int \left(\hat{G}_i(t) - G(t)\right)^2 dt,$$

sendo G(t) frequentemente substituído na avaliação por

$$\overline{G}_i(t) = \frac{\sum_{j \neq i} \hat{G}_j(t)}{s-1}.$$

Aplicações

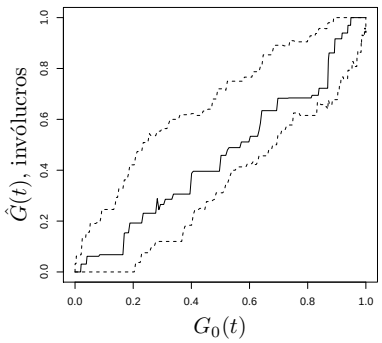


Figura 4.15 *Distribuição empírica das distâncias ao vizinho mais próximo e invólucros contra a correspondente distribuição teórica, sob a hipótese de aleatoriedade espacial completa.*

Na Figura 4.15 encontra-se, relativamente ao exemplo das mudas de pinheiro negro Japonês, a função distribuição empírica das distâncias ao vizinho mais próximo contra a distribuição teórica desta função sob a validade da hipótese de aleatoriedade espacial completa (a cheio) e os correspondentes invólucros superior e inferior calculados com base em 99 simulações da aleatoriedade espacial completa (a tracejado).

Este gráfico sugere a aceitação desta hipótese, o que é confirmado pelo teste Monte Carlo efectuado com base nos postos dos u_i acima definidos (valor p do teste igual 0.58).

Seguidamente mostramos na Figura 4.16 o mesmo tipo de gráfico, mas agora para os conjuntos de dados dos rebentos de Pau-Brasil (esquerda) e das localizações dos incêndios em Portugal Continental em 1975 (direita).

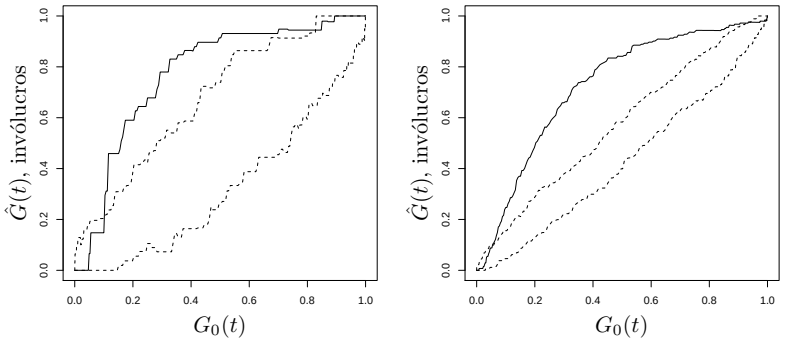


Figura 4.16 *Distribuição empírica das distâncias ao vizinho mais próximo e invólucros contra a correspondente distribuição teórica sob a hipótese de aleatoriedade espacial completa, para os dados dos rebentos do Pau-Brasil (esquerda) e os dados das localizações dos incêndios (direita).*

Verifica-se em ambos os casos que os gráficos apontam no sentido da rejeição da hipótese de aleatoriedade espacial completa, o que é confirmado pelos testes de Monte Carlo que, em ambos os casos, resultam num valor p de 0.01.

Teste usando distâncias entre um ponto e o acontecimento mais próximo

Por analogia com o procedimento empregue para as distância ao vizinho mais próximo, um teste Monte Carlo para testar a hipótese de aleatoriedade espacial assenta em $F(t)$ e na sua distribuição empírica, dada por (4.6) ou (4.7), e é baseado na estatística

$$u_i = \int \left(\hat{F}_i(t) - \overline{F}_i(t)\right)^2 dt,$$

com $\hat{F}_i(t)$ dado por

$$\overline{F}_i(t) = \frac{\sum_{j \neq i} \hat{F}_j(t)}{s-1}.$$

Aplicações

Confirmamos agora, com recurso às funções distribuição de um ponto ao acontecimento mais próximo, as constatações atrás efectuadas relativas à hipótese de aleatoriedade espacial completa feitas para os conjuntos de dados que temos vindo a considerar. Os testes Monte Carlo realizados agora resultam numa aceitação desta hipótese nos casos dos dados das mudas de pinheiro negro japonês e numa rejeição veemente nos outros dois casos. Nas Figuras 4.17 e 4.18 apresentam-se os gráficos das distribuições empíricas e invólucros das distâncias de um ponto ao acontecimento mais próximo contra a correspondente distribuição sob a hipótese de aleatoriedade espacial completa, para os três conjuntos de dados.

Verifica-se em ambos os casos que os gráficos apontam no sentido da rejeição da hipótese de aleatoriedade espacial completa, o que é

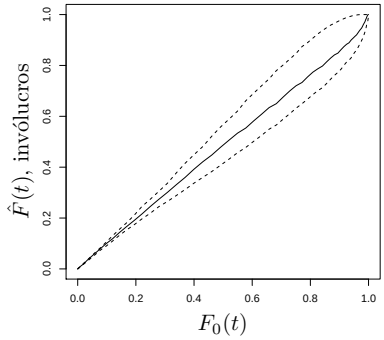


Figura 4.17 *Distribuição empírica das distâncias de um ponto ao vizinho mais próximo e invólucros contra a correspondente distribuição teórica sob a hipótese de aleatoriedade espacial completa, para o exemplo das mudas de pinheiro negro Japonês.*

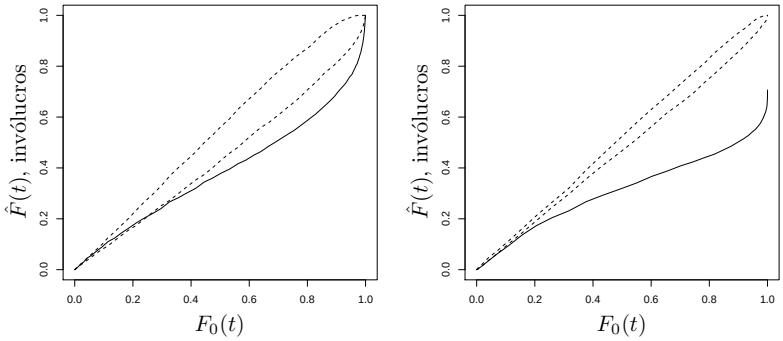


Figura 4.18 *Distribuição empírica das distâncias de um ponto ao vizinho mais próximo e invólucros contra a correspondente distribuição teórica sob a hipótese de aleatoriedade espacial completa, para os dados dos rebentos do Pau-Brasil (esquerda) e os dados das localizações dos incêndios (direita).*

confirmado pelos testes de Monte Carlo que, em ambos os casos, resultam num valor p de 0.01.

Teste usando invólucros baseados na função K

De forma análoga à descrita anteriormente para as distribuições dos vizinhos, é possível determinar invólucros superiores e inferiores a partir de simulações da aleatoriedade espacial completa para a função K. Se a função K empírica se encontrar dentro dos dois invólucros, não se rejeita esta hipótese.

A função `invólucro` do *package* `spatstat` do programa `R` encarrega-se do calculo destes invólucros para todas as medidas resumo.

Teste à aglomeração usando as funções K

Frequentemente, em epidemiologia, está-se interessado na localização da incidência de uma doença na população. No entanto, os testes simples à hipótese de aleatoriedade espacial completa não são, geralmente para estes casos, apropriados, por causa da enorme falta de homogeneidade na população.

Uma forma adequada de se conseguir comparar a distribuição dos casos com a população em risco é obter-se um conjunto de pontos tirados como uma amostra aleatória da população em risco e usar estes pontos como pontos controle, para comparação com os dados. Podemos então determinar estimativas das funções K para os dados e para os controles, comparando-as posteriormente.

4.4 Modelos

A pedra basilar da modelação dos processos pontuais é o processo de Poisson. O processo de Poisson homogéneo traduz a situação

de aleatoriedade espacial completa, onde os acontecimentos se posicionam de forma independente uns dos outros.

Quando a ocorrência de um acontecimento num determinado local torna mais provável a ocorrência de um outro acontecimento próximo, o padrão resultante pode ser descrito como sendo agregado. Se, pelo contrário, cada acontecimento é mais provável de estar rodeado de um espaço vazio, o padrão resultante será de uma distribuição de acontecimentos mais regularmente distribuídos. Nesta secção apresentaremos os modelos adequados à descrição destes e outros padrões possíveis.

Para estes dados, modelos realísticos devem incorporar tanto a não-homogeneidade espacial ('tendência') como dependência entre pontos ('interacção' tal como a aglomeração ou a regularidade) - Baddeley e Turner (2000).

Os processos espaciais pontuais podem ser especificados a partir de uma função intensidade determinística ou aleatória, por analogia com os modelos lineares generalizados. Há em particular duas classes de modelos importantes que providenciam modelos sem interacção (processo de Poisson) e padrões pontuais agregados (processos agregados e de Cox).

4.4.1 Processos de Poisson agregados

Estes processos descrevem padrões pontuais espaciais agregados resultantes da aglomeração de grupos de acontecimentos relacionados.

Definição 4.14. *Um processo pontual diz-se um* ***Processo de Poisson agregado*** *se:*

PA1 *Acontecimentos pais formam um processo de Poisson (homogéneo) com intensidade ρ.*

PA2 *Cada pai produz, independentemente e com idêntica distribuição de probabilidade* $p_S, \quad S = 0, 1, \ldots,$ *um número aleatório de descendentes* S.

PA3 *As posições dos descendentes em relação aos seus pais são independentes e identicamente distribuídas de acordo com uma função densidade probabilidade bivariada* $h(\cdot)$.

Convenciona-se aqui que o padrão final resulta apenas da descendência, o que não é unanimemente considerado por todos os autores.

Proposição 4.9. *Um processo de Poisson agregado satisfaz as seguintes propriedades:*

1. *É um processo estacionário com intensidade* $\lambda = \rho\mu$, *onde* $\mu = \mathrm{E}[S]$.
2. *É um processo isotrópico se* ***PA3*** *especifica uma função densidade probabilidade radialmente simétrica.*
3. $$K(t) = \pi t^2 + \frac{\mathrm{E}[S(S-1)]H_2(t)}{\rho\mu^2}, \tag{4.10}$$

   onde $H_2(t)$ *é a função distribuição do vector diferença entre as posições de dois descendentes do mesmo pai.*
4. $$\lambda_2(t) = \lambda^2 + \rho\mathrm{E}[S(S-1)]h_2(t),$$

   onde $h_2(t)$ *é a função densidade de probabilidade do vector diferença apresentado no ponto anterior.*

Note-se que de (4.10) se vê que:

- O seu segundo termo é não negativo e monótono não decrescente;

- $\lim_{t\to\infty} \left\{K(t) - \pi t^2\right\} = \dfrac{\mathrm{E}[S(S-1)]}{\rho\mu^2} = c$, constante;
- Se S seguir uma distribuição Poisson , $c = \frac{1}{\rho}$;
- As observações anteriores sugerem uma forma útil de identificar se um processo de Poisson agregado pode ser um modelo razoável para um padrão observado e, caso afirmativo, meios para obter estimativas preliminares dos parâmetros.

Os aglomerados obtidos nos processos pontuais não homogéneos provocados pela heterogeneidade ambiental estão por oposição aos obtidos aqui com os processos de Poisson agregados.

Apresentamos na Figura 4.19 uma realização de um processo de Poisson agregado com 20 pais em média por unidade quadrado e 5 descendentes em média por cada pai, com um raio de cada aglomerado não excedendo 0.05 (esta condição é forçada via função $h(\cdot)$) - simulado com a função `rMatClust` do `spatstat`.

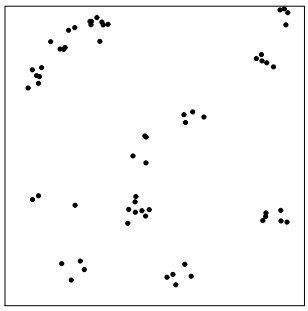


Figura 4.19 *Realização de um processo de Poisson agregado.*

4.4.2 Processos de Cox

A fonte da heterogeneidade ambiental atrás mencionada pode também ser estocástica por natureza. A incorporação desta observação

nos processos de Poisson não homogéneos atrás descritos dão origem aos processos de Cox, uma classe de modelos formados como processos de Poisson não homogéneos com funções intensidade estocásticas.

Definição 4.15. *Um* ***Processo de Cox*** *define-se como aquele que satisfaz o seguinte:*

PC1 $\{\Lambda(x) : x \in \mathbb{R}^2\}$ *é um processo estocástico tomando apenas valores não negativos.*

PC2 *Condicional a uma realização de* $\Lambda(x)$, $\{\Lambda(x) = \lambda(x) : x \in \mathbb{R}^2\}$, *os acontecimentos formam um processo de Poisson não homogéneo com função intensidade* $\lambda(x)$.

Costuma-se dizer que o processo de Cox é derivado de $\{\Lambda(x)\}$.

Proposição 4.10. *Um processo de Cox satisfaz as seguintes propriedades:*

1. *É estacionário se e só se o processo das intensidades* $\Lambda(x)$ *é estacionário.*
2. *É isotrópico se e só se o processo das intensidades* $\Lambda(x)$ *é isotrópico.*
3. *Sendo estacionário a intensidade é* $\lambda = \mathrm{E}[\Lambda(x)]$.
4. $\lambda_2(x, y) = \mathrm{E}[\Lambda(x)\Lambda(y)]$.
5. *Havendo estacionaridade e isotropia,* $\lambda_2(t) = \lambda^2 + \vartheta(t)$, *onde* $\vartheta(t) = \mathrm{Cov}(\Lambda(x), \Lambda(y))$ *e* $t = \|x - y\|$.
6. $K(t)$ *sai de (4.2),*
$$\lambda K(t) = \frac{2\pi}{\lambda} \int_0^t \lambda_2(x) x \, dx.$$
7. *No caso típico em que* $\vartheta(t)$ *toma apenas valores não negativos a expressão de* $\lambda_2(t)$ *acima é semelhante à correspondente expressão no processo de Poisson agregado, pela equivalência dos processos nas duas classes.*

Como consequência do último ponto, de uma perspectiva estatística, a distinção entre aglomeração e heterogeneidade apenas pode ser sustentada se houver informação adicional, por exemplo sob a forma de covariáveis. Note-se que se conseguíssemos modelar a superfície da intensidade $\Lambda(x)$ através de uma equação de regressão com covariáveis, em vez de ser uma realização de um processo estocástico, o modelo de processo de Poisson resultante viria um processo de Poisson não homogéneo.

Uma construção relativamente flexível e matematicamente manipulável para os processos de Cox, não baseada na sua dualidade com os processos agregados, é a classe de processos log-gaussianos. Aí modelamos o logaritmo do processo intensidade $\Lambda(x)$ com recurso a covariáveis mais um erro que é um processo gaussiano - obtemos um processo de Cox log-gaussiano.

4.4.3 Processos de inibição simples

Estes processos surgem quando se pensa modelar padrões mais regulares, imaginando-se que o desvio do padrão de aleatoriedade espacial completa acontece apenas porque se impõe a existência de uma distância admissível mínima, δ, entre quaisquer dois acontecimentos. Por exemplo tal acontece quando temos árvores a competir entre si por luz.

Definição 4.16. *Define-se intensidade de empacotamento (packing intensity) de um processo de inibição simples como a proporção do plano coberto por discos não sobrepostos de diâmetro δ ou como a proporção esperada de cobertura de uma região finita A:*

$$\tau = \lambda\pi\delta^2/4,$$

onde λ é a intensidade.

Note-se que a intensidade de empacotamento máxima é obtida por discos colados cujos centros formam um reticulado triangular equilateral com espaçamento δ. Consequentemente,

Proposição 4.11. $\tau \leq \tau_{max} = \dfrac{\sqrt{3}\pi}{6} \simeq 0.91.$

Enquanto que um processo de inibição simples corresponde a um processo de Poisson adelgaçado por se apagar todos os pares de eventos que distem menos do que δ, um processo de inibição sequencial simples marca cada acontecimento com o seu tempo de nascimento e só remove os acontecimentos que distem menos de δ de um acontecimento mais antigo.

Definição 4.17. *Define-se um* ***Processo de inibição sequencial simples*** *numa região finita A como uma sequência de acontecimentos X_i em A tais que:*

IS1 *X_1 segue uma distribuição Uniforme em A.*

IS2 *Dado $\{X_j = x_j,\ j = 1, \ldots, i-1\}$, X_i segue uma distribuição Uniforme na intersecção $A \cap \{y : \| y - x_j \| \geq \delta,\ j = 1, \ldots, i-1\}$.*

Estes processos são muito naturalmente parametrizados pela respectiva intensidade de empacotamento, $\tau = n\pi\delta^2/4|A|$.

Note-se que se se prescrever um grande valor de τ o procedimento sequencial pode terminar prematuramente. O máximo atingido por esta intensidade é uma variável aleatória com uma distribuição perfeitamente intratável, mas para o qual estudos de simulação sugerem um valor esperado de 0.547.

Proposição 4.12. *Para um processo de inibição simples temos que:*

1.

$$\lambda = \frac{1 - \exp(-\pi\rho\delta^2)}{\pi\delta^2}$$

2.

$$\lambda_2(t) = \begin{cases} 0, & t < \delta \\ \frac{2U_\delta(t)\left(1-\exp(-\rho\pi\delta^2)\right)-2\pi\delta^2\left(1-\exp(-\rho U_\delta(t))\right)}{\rho^2\pi\delta^2 U_\delta(t)\left(U_\delta(t)-\pi\delta^2\right)}, & t \geq \delta \end{cases},$$

onde $U_\delta(t)$ denota a área da união de dois discos cada com raio δ e com centros distantes um do outro de t.

Mostra-se na Figura 4.20 uma realização de um processo de inibição simples, numa janela de lado unitário, com distância mínima entre acontecimentos de $\delta = 0.1$.

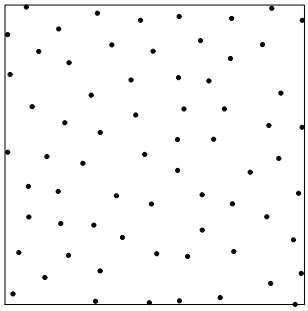


Figura 4.20 *Realização de um processo de inibição simples.*

4.4.4 Processos pontuais de Markov

A motivação para esta classe de modelos tem a ver com o se pretender preservar a dependência local markoviana entre acontecimentos da classe de modelos de inibição, mas incluir um pouco mais de flexibilidade na formulação. Servem para modelar tanto padrões agregados como regulares.

Os processos pontuais de Markov são definidos para uma região finita arbitrária, mas fixa, A. Todos os modelos que não o processo de Poisson homogéneo, caracterizam-se pela sua razão de verosimilhanças $f(\cdot)$ com respeito a um processo de Poisson de intensidade unitária. Assume-se pois, que o processo pontual tem função densidade probabilidade $f(\cdot)$ com respeito à distribuição do processo de Poisson com intensidade 1 em A.

Assim, se $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ denota um conjunto finito de pontos em A, $f(\boldsymbol{x})$ indica, intuitivamente, quão mais provável é a configuração de acontecimentos $\boldsymbol{x}$ para o processo particular do que para um processo de Poisson com intensidade unitária. Por exemplo, o processo de Poisson homogéneo com intensidade λ tem densidade:

$$f(\boldsymbol{x}) = \exp(-(\lambda - 1)|A|)\lambda^n,$$

e um processo de Poisson não homogéneo em A com intensidade $\lambda(x)$ tem densidade dada por:

$$f(\boldsymbol{x}) = \exp\left(-\int_A \big(\lambda(u) - 1\big)\, du\right) \prod_{i=1}^{n} \lambda(x_i).$$

Adicionalmente assumimos que $f(\boldsymbol{x}) > 0$ implica $f(\boldsymbol{y}) > 0$ para todos os subconjuntos $\boldsymbol{y} \subset \boldsymbol{x}$ - classe dos **processos de Gibbs** em A.

Usualmente, $f(\cdot)$ é definido a menos de uma constante normalizadora que não pode ser determinada de forma explícita. Note-se, contudo, que $f(\cdot)$ pode sempre ser factorizada como:

$$f(\boldsymbol{x}) = c \prod_{i=1}^{n} g_i(x_i) \prod_{j>i} g_{ij}(x_i, x_j) \cdot \ldots \cdot g_{12\ldots n}(x_1, \ldots, x_n). \quad (4.11)$$

Definição 4.18. *Dizemos que os pontos x e y em A são* ***vizinhos*** *se $\|x - y\| < \rho$, onde $\rho > 0$ é um valor escolhido à partida (****alcance do***

processo). *Definimos ainda* ***vizinhança*** *de um ponto* $x \in A$ *como o conjunto de pontos* $\{y \in A : 0 < \|x - y\| < \rho\}$ *e definimos uma* ***clique*** *como um conjunto de vizinhos mútuos.*

Definição 4.19. *Um* ***Processo pontual*** *é dito de* ***Markov de alcance*** ρ *se a intensidade condicional no ponto* x*, dada a configuração de acontecimentos no restante conjunto* A*, depende apenas da configurações dos pontos na vizinhança de* x*.*

Assim, adicionando a um processo de Gibbs uma condição de vizinho mais próximo, resulta num processo de Markov.

Para um processo Pontual ser de Markov de alcance ρ é necessário que cada função g em (4.11) seja unitária a menos que os seus argumentos formem uma clique. Outras condições devem ser impostas, essencialmente para que $f(\cdot)$ seja integrável.

Esta formulação dos processos Pontuais de Markov encontra-se estreitamente relacionada com o Teorema de factorização de Hammersley-Clifford (Besag, 1974).

Definição 4.20. *Um* ***Processo de Strauss*** *é um exemplo de um processo pontual de Markov para o qual:*

$$f(\boldsymbol{x}) = c\beta^n\gamma^s,$$

onde n *é o número de acontecimentos em* $\boldsymbol{x}$*,* s *é o número de pares distintos de vizinhos,* c *é a constante normalizadora,* β *reflecte a intensidade do processo e* γ *descreve as interacções entre vizinhos.*

O caso $\gamma = 1$ é um processo de Poisson com intensidade β e o caso $\gamma = 0$ é um processo de inibição simples. Para valores de γ entre 0 e 1 há inibição entre pares de pontos próximos (inibição não estrita). Quando $gamma > 1$ o processo explode, com um número

de acontecimentos infinitos em A. Se se fixar o n, tal resulta numa distribuição de probabilidade válida para $\boldsymbol{x}$ para qualquer $\gamma > 0$. No entanto, para $\gamma > 1$, o padrão resultante tende a exibir uma forma extrema de aglomeração.

O processo de Strauss é um caso particular do seguinte:

Definição 4.21. *Define-se a classe dos* ***processos de interacção emparelhada*** *como:*

$$f(\boldsymbol{x}) = c\beta^n \prod_i \prod_{j \neq i} h(\|x_i - x_j\|, \boldsymbol{\phi}), \tag{4.12}$$

onde $h(u, \boldsymbol{\phi})$ é uma função não-negativa que pode depender de parâmetros $\boldsymbol{\phi}$ a serem estimados dos dados. De forma a se obter um processo pontual válido tem de se restringir $h(\cdot)$, por exemplo, escolhendo-a limitada e de tal forma que $h(u; \boldsymbol{\phi}) = 0,\ u < \delta,\ \delta > 0$.

O termo β pode não ser constante, influenciando nesse caso a intensidade dos pontos e introduzindo tendência espacial, $\beta(x_i; \boldsymbol{\phi})$. Os termos $h(\|x_i - x_j\|, \boldsymbol{\phi})$ reflectem dependência ou interacção entre diferentes pontos do processo X. Caso esta função seja unitária o modelo reduz-se ao processo de Poisson homogéneo com intensidade β e, no caso de este não ser constante, ao processo de Poisson não homogéneo com intensidade $\beta(x; \boldsymbol{\phi})$.

Existem outros casos particulares destes processos de interacção emparelhados, como por exemplo o processo pontual dos triplos (Ripley e Kelly, 1977), em que a ideia é adicionar a clique da próxima maior ordem para obter um processo que permita a atracção positiva de pares de pontos e outros, imprimindo um maior ou menor grau de dependência entre os acontecimentos.

4.5 Ajustamento de modelos

Os métodos clássicos de inferência para os processos pontuais são muito condicionados pelo facto de as funções de verosimilhança da maioria dos modelos de interesse ter um tratamento matemático extremamente difícil. A abordagem numérica é computacionalmente muito pesada e emprega algoritmos de simulação que são específicos do modelo escolhido.

Como alternativa podemos usar métodos baseados na comparação de medidas resumo teóricas com as mesmas medidas empíricas, o que constitui uma forma simples e rápida de se conseguir explorar uma grande diversidade de modelos.

Naturalmente que temos ainda outra opção, se nos quisermos libertar da restrição de trabalhar com famílias de modelos paramétricas, que é recorrer a métodos não paramétricos para fazer inferência - Diggle(2003).

4.5.1 Ajustamento de modelos usando medidas descritivas

A grande vantagem destes métodos, quando comparados com os métodos baseados na máxima verosimilhança, é a sua grande facilidade de implementação na prática.

Usando $K(t)$

A ideia de se estimar a função $K(t)$ surge por esta função ter uma forma conhecida para um grande número de modelos diferentes de processos pontuais. A forma observada de $\hat{K}(t)$ pode sugerir modelos candidatos e estimativas iniciais dos respectivos modelos.

Represente $K(t;\boldsymbol{\theta})$ a função teórica do nosso modelo, que depende dos parâmetros $\boldsymbol{\theta}$, e seja $\hat{K}(t)$ o estimador desta função, dado por (4.5).

De forma a medir-se a discrepância entre a função $K(t)$ teórica e o seu estimador, define-se a seguinte classe de funções:

$$D(\boldsymbol{\theta}) = \int_0^{t_0} w(t)\Big(\big(\hat{K}(t)\big)^c - \big(K(t;\boldsymbol{\theta})\big)^c\Big)^2 dt,$$

onde as constantes t_0 e c e a função dos pesos $w(t)$ são a escolher.

Estimamos então, pelo método dos mínimos quadrados, $\boldsymbol{\theta}$, como aquele valor $\hat{\boldsymbol{\theta}}$ que minimiza $D(\boldsymbol{\theta})$.

A escolha dos valores das constantes t_0 e c e dos pesos w prende-se com considerações de cariz essencialmente empírico, notando porém que a sua escolha se reflecte nas flutuações amostrais do estimador $\hat{K}(t)$.

Assim é sugerido que $c = 0.5$ em conjunção com $w(t) = 1$ é uma escolha razoável para ajustar modelos a padrões regulares, mas para padrões agregados uma escolha de $c = 0.25$ é frequentemente mais eficaz. Para dados no quadrado unitário, recomenda-se que t_0 não exceda o valor 0.25 (e proporcionalmente para áreas maiores), sendo que a escolha de t_0 muito pequeno se concentra nos efeitos de pequena escala e reciprocamente. Será bom experimentar vários valores de t_0 e c, de modo a avaliar de que forma os resultados são sensíveis às suas escolhas.

Alternativamente podemos fixar $c = 1$ e usar pesos inversamente proporcionais à variância amostral de $\hat{K}(t)$ - para padrões próximos do de Poisson, $w(t) = \frac{1}{t^2}$.

A distribuição amostral de $\hat{\boldsymbol{\theta}}$ poderá ser obtida pelo método Monte Carlo, ainda que tal possa ser bastante pesado computacionalmente.

Se $K(t;\boldsymbol{\theta})$ não puder ser avaliado quer explicitamente quer numeri-

camente o que se pode fazer é substituí-lo em $D(\boldsymbol{\theta})$ atrás pela média de s suas estimativas pontuais obtidas da simulação de s realizações do modelo teórico pressuposto.

Bondade do ajustamento via distribuições de vizinho mais próximo

Para avaliar a qualidade do ajuste podemos recorrer a duas estatísticas de bondade de ajustamento baseadas das funções distribuição empíricas das distâncias do vizinho mais próximo e de um ponto ao acontecimento mais próximo, à semelhança do que já fizemos quando pretendemos testar a hipótese de aleatoriedade completa:

$$g_i = \int_0^\infty \left(\hat{G}_i(t) - \overline{G}_i(t)\right)^2 dt$$
$$f_i = \int_0^\infty \left(\hat{F}_i(t) - \overline{F}_i(t)\right)^2 dt,$$

onde $\hat{G}_1(t)$ é a função distribuição empírica da distância do vizinho mais próximo calculada para os dados, $\hat{G}_i(t),\ i = 2, \ldots, s$ são as funções distribuição empírica da distância do vizinho mais próximo calculadas a partir de simulações do modelo pressuposto e

$$\overline{G}_i(t) = \frac{\sum_{j \neq i}^{s} \hat{G}_j(t)}{s-1}.$$

(Semelhantes definições para $\hat{F}_i(t)$ e $\overline{F}_i(t)$).

Obtendo a significância dos dois testes, digamos p_1 e p_2, sabemos que a significância global p se encontra entre $\min\{p_1, p_2\}$ e $2\min\{p_1, p_2\}$ - Cox(1977).

Estimação paramétrica através de testes da bondade do ajustamento

Imaginemos que temos um modelo dependente de um conjunto de parâmetros em $\boldsymbol{\theta}$, desconhecidos, e um conjunto de dados que se assumem gerados por tal modelo. Para cada valor fixo de $\boldsymbol{\theta}$ podemos testar a bondade do ajustamento do modelo. Então, o conjunto de valores de $\boldsymbol{\theta}$ que não são rejeitados por teste a um nível de significância α constituem uma região de confiança $(1-\alpha)\times 100\%$ para $\boldsymbol{\theta}$.

Se os testes individuais forem testes de Monte Carlo, claro que esta ideia involve uma discretização do espaço paramétrico, apenas admissível no caso de $\boldsymbol{\theta}$ ser de pequena dimensão.

4.5.2 Ajustamento de modelos baseada na máxima verosimilhança

A estimação via verosimilhança é um verdadeiro desafio nos processos pontuais, dado que na grande maioria dos casos as funções de verosimilhança são intratáveis. Esta dificuldade tem sido vencida nos últimos anos pelos desenvolvimentos dos métodos de estimação baseados nos métodos de Cadeias de Markov de Monte Carlo.

No processo de Poisson não homogéneo com função intensidade $\lambda(x)$, contudo, a verosimilhança é tratável - a distribuição associada com uma realização parcial $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ deste processo numa região finita A pode ser factorizada como o produto de uma distribuição Poisson com média $\mu = \int_A \lambda(x)\,dx$ para o número de acontecimentos n, e um conjunto de localizações independentes x_i cuja distribuição comum tem densidade $\frac{\lambda(x)}{\mu}$,

$$\frac{\exp(-\mu)\mu^n}{n!} \times \prod_{i=1}^{n} \frac{\lambda(x_i)}{\mu},$$

resultando na seguinte log-verosimilhança:

$$L(\lambda) = \sum_{i=1}^{n} \log(\lambda(x_i)) - \int_A \lambda(x)\, dx.$$

Na prática a utilização da verosimilhança é mais útil se $\lambda(x)$ puder ser especificado através de um modelo de regressão, por exemplo um modelo log-linear:

$$\log \lambda(x) = \sum_{j=1}^{p} \beta_j z_j(x),$$

sendo $z_j(x)$ o valor das covariáveis em x.

O integral na log-verosimilhança acima pode ser um problema, exigindo que as covariáveis z_j sejam medidas continuamente sobre toda a região em estudo. Berman e Turner (1992) contornam esta dificuldade, aproximando a log-verosimilhança por uma soma finita com a mesma forma analítica que a log-verosimilhança (ponderada) de um modelo linear generalizado com respostas Poisson, usando então *software* já desenvolvido para este tipo de modelos - implementado também no pacote `spatstat` do `R`.

Consideremos agora o caso dos processos pontuais de Markov, definidos com base na distribuição conjunta $f(\boldsymbol{x})$ de uma configuração de pontos em A, $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, dada por (4.11), e que serve de base à inferência por máxima verosimilhança (n fixo). O grande problema é que a constante normalizadora, c, não se consegue obter analiticamente, na maioria dos casos.

Inferência por pseudo-verosimilhança

Uma forma de ultrapassar este problema foi encontrada por Besag(1975) e Besag *et al.*(1982) e reavivada no uso por Baddeley e Turner (2000), numa generalização do procedimento de Berman e Turner (1992) atrás referido. Consiste em substituir na inferência

o uso da função de verosimilhança por uma outra função chamada **função de pseudoverosimilhança**, definida, para uma distribuição multivariada geral $f(x_1, \ldots, x_n)$, como o produto das chamadas **condicionais completas**. Segue-se a função de log-pseudo-verosimilhança:

$$PL = \sum_{i=1}^{n} \log\big(f(x_i|x_j, j \neq i)\big).$$

A intensidade condicional de um ponto localizado arbitrariamente em x, dada a realização $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ do processo em A é:

$$\lambda(x; \boldsymbol{x}) = \begin{cases} \frac{f\left(\boldsymbol{x} \cup \{x\}\right)}{f(\boldsymbol{x})}, & x \notin \boldsymbol{x} \\ \\ \frac{f(\boldsymbol{x})}{f\left(\boldsymbol{x} \setminus \{x\}\right)}, & x \in \boldsymbol{x} \end{cases},$$

resultando na seguinte função log-pseudoverosimilhança:

$$PL_\lambda = \sum \log\big(\lambda(x_i; \boldsymbol{x})\big) - \int_A \lambda(x; \boldsymbol{x})\, dx.$$

Assim, para a classe de processos de interacção emparelhada definidos por (4.12), temos que:

$$\lambda(x; \boldsymbol{x}) = \begin{cases} \beta \prod_{j=1}^{n} h(\|x - x_j\|, \boldsymbol{\phi}), & x \notin \boldsymbol{x} \\ \\ \beta \prod_{j \neq i}^{n} h(\|x_i - x_j\|, \boldsymbol{\phi}), & x = x_i \in \boldsymbol{x} \end{cases},$$

tendo desaparecido a constante normalizadora da função densidade e resultando na seguinte função log-pseudoverosimilhança:

$$\begin{aligned} PL(\beta, \boldsymbol{\phi}) = n \log(\beta) &+ \sum_{i=1}^{n} \sum_{j \neq i} \log\Big(h(\|x_i - x_j\|; \boldsymbol{\phi})\Big) \\ &- \beta \int_A \exp\left\{\sum_{j=1}^{n} \log\Big(h(\|x - x_j\|; \boldsymbol{\phi})\Big)\right\} dx. \end{aligned}$$

O estimador de pseudoverosimilhança para β sai directamente da maximização da função acima,

$$\hat{\beta}_h = \frac{n}{\int_A \exp\left\{\sum_{j=1}^{n} \log\left(h(\|x - x_j\|; \boldsymbol{\phi})\right)\right\} dx}$$

Substituindo $\hat{\beta}_h$ na log-pseudoverosimilhança acima obtemos a função de log-pseudoverosimilhança reduzida:

$$PL_0(\phi) = n(\log n - 1) + \sum_{i=1}^{n} \sum_{j \neq i} \log\left(h(\|x_i - x_j\|; \boldsymbol{\phi})\right) - \\ - n \log\left(\int_A \exp\left\{\sum_{j=1}^{n} \log\left(h(\|x - x_j\|; \boldsymbol{\phi})\right)\right\} dx\right)$$

Maximiza-se numericamente esta função para obter-se o estimador de máxima peuso-verosimilhança $\hat{\boldsymbol{\phi}}$, que depois se substitui em $\hat{\beta}_h$ para se obter o estimador de máxima pseudoverosimilhança, $\hat{\beta}$.

Os estimadores de máxima pseudoverosimilhança são então uma alternativa prática à estimação por máxima verosimilhança, sendo consistentes e assimptoticamente normais sob condições adequadas. Permite, com a generalização do dispositivo de Berman e Turner (1992) e *software standard*, ajustar rapidamente modelos estocásticos espaciais complexos envolvendo tendências, covariáveis espaciais e interacções entre pontos.

A pseudoverosimilhança é a base do algoritmo de Baddeley e Turner (2000) para ajustar modelos de processos pontuais muito genéricos, estando implementado no pacote do `R` `spatstat` (Baddeley e Turner, 2005, 2006), e sendo de utilização muito atractiva.

É possível incorporar correcções dos efeitos fronteira nesta estimação por verosimilhança penalizada, ponderando na função $PL_0(\cdot)$

os termos de $\log\big(h(\cdot)\big)$ com os mesmos pesos iguais aos que usámos na estimação da função K - Diggle *et al.* (1994).

Nas aplicações práticas será útil ter um estimador simples e não paramétrico da função interacção $h(\cdot)$. Uma solução é a de usar pseudoverosimilhança em conjunção com uma especificação de $h(\cdot)$ constante por troços, levando a uma espécie de estimador histograma desta função - Baddeley e Turner (2000).

Inferência Monte Carlo

Alternativamente pode fazer-se inferência recorrendo a métodos Monte Carlo para ultrapassar a dificuldade do tratamento matemático da verosimilhança.

Para a classe de processos de interacção emparelhada definidos por (4.12), representando $\boldsymbol{\theta} = (\beta, \boldsymbol{\phi})$ os parâmetros do modelo, temos a seguinte verosimilhança:

$$\ell(\boldsymbol{\theta}) = c(\boldsymbol{\theta}) f^*(\boldsymbol{x}; \boldsymbol{\theta})$$

onde $f^*(\boldsymbol{x}; \boldsymbol{\theta}) = \prod_i \prod_{j \neq i} h(\|x_i - x_j\|; \boldsymbol{\theta})$. O problema é a determinação da constante normalizadora $c(\boldsymbol{\theta})$, que naturalmente é dada por:

$$\begin{aligned} c(\boldsymbol{\theta})^{-1} &= \int_{A^n} f^*(\boldsymbol{x}; \boldsymbol{\theta}) = \\ &= \int_{A^n} f^*(\boldsymbol{x}; \boldsymbol{\theta}) \times \frac{c(\boldsymbol{\theta}_0)}{c(\boldsymbol{\theta}_0)} \times \frac{f^*(\boldsymbol{x}; \boldsymbol{\theta}_0)}{f^*(\boldsymbol{x}; \boldsymbol{\theta}_0)}, \end{aligned}$$

sendo $\boldsymbol{\theta}_0$ um qualquer ponto no espaço paramétrico.

Definindo

$$r(\boldsymbol{x}; \boldsymbol{\theta}, \boldsymbol{\theta}_0) = \frac{f^*(\boldsymbol{x}; \boldsymbol{\theta})}{f^*(\boldsymbol{x}; \boldsymbol{\theta}_0)},$$

a constante ainda se pode escrever como:

$$\begin{aligned} c(\boldsymbol{\theta})^{-1} &= c(\boldsymbol{\theta}_0)^{-1} \int_{A^n} r(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\theta}_0) c(\boldsymbol{\theta}_0) f^*(\boldsymbol{x};\boldsymbol{\theta}_0)\, d\boldsymbol{x} \\ &= c(\boldsymbol{\theta}_0)^{-1} \mathrm{E}_{\boldsymbol{\theta}_0}\big[r(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\theta}_0)\big], \end{aligned}$$

sendo $\mathrm{E}_{\boldsymbol{\theta}_0}$ o valor esperado com respeito à distribuição de $\boldsymbol{x}$, quando $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Com este resultado, se conseguirmos encontrar um valor de $\boldsymbol{\theta}_0$ com o qual consigamos avaliar $c(\boldsymbol{\theta}_0)$ directamente e simular realizações $\boldsymbol{x}$, então poderemos avaliar $c(\boldsymbol{\theta})$ aproximadamente para qualquer valor de $\boldsymbol{\theta}$, usando simulações $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s$ sob $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ e obtendo uma aproximação empírica da esperança:

$$\hat{\mathrm{E}}_{\boldsymbol{\theta}_0}\big[r(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\theta}_0)\big] = \frac{\sum_{j=1}^{s} r(\boldsymbol{x}_j,\boldsymbol{\theta},\boldsymbol{\theta}_0)}{s}$$

Note-se que este método pode falhar quando o valor de $\boldsymbol{\theta}$ não se encontra próximo de $\boldsymbol{\theta}_0$, porque a razão em $r(\cdot)$ é instável.

Temos ainda que a verosimilhança se pode escrever como:

$$\ell(\boldsymbol{\theta}) = \frac{c(\boldsymbol{\theta}_0) f^*(\boldsymbol{x};\boldsymbol{\theta})}{\mathrm{E}_{\boldsymbol{\theta}_0}\big[r(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\theta}_0)\big]}.$$

Daqui podemos pensar numa forma de obter o estimador de máxima verosimilhança $\hat{\boldsymbol{\theta}}$ aproximado. Sendo $\boldsymbol{\theta}_0$ constante, para qualquer $\boldsymbol{\theta}_0$, o estimador de máxima verosimilhança deve maximizar:

$$L_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}) = \log\big(f^*(\boldsymbol{x};\boldsymbol{\theta})\big) - \log\Big(\mathrm{E}_{\boldsymbol{\theta}_0}\big[r(\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\theta}_0)\big]\Big).$$

Assim, a aproximação que procuramos pode ser obtida da maximização de

$$\hat{L}_{\boldsymbol{\theta}_0,s}(\boldsymbol{\theta}) = \log\big(f^*(\boldsymbol{x};\boldsymbol{\theta})\big) - \log\left(\frac{\sum_{j=1}^{s} r(\boldsymbol{x}_j;\boldsymbol{\theta},\boldsymbol{\theta}_0)}{s}\right),$$

onde $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s$ são realizações simuladas com $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Os comentários atrás sobre a admissibilidade do uso da função $r(\cdot)$ para avaliar a constante normalizadora $c(\boldsymbol{\theta})$ continuam aqui a ser pertinentes. No entanto podemos recorrer à máxima pseudo verosimilhança penalizada para escolher um valor para $\boldsymbol{\theta}_0$, que podemos considerar como uma primeira aproximação à estimação de máxima verosimilhança.

4.5.3 Aplicações

Neste curso, pela sua brevidade, não é possível detalhar-se exaustivamente de todos os modelos de processos pontuais à nossa disposição. Como tal, a parte da selecção, ajustamento e validação de um modelo aos conjuntos de dados que temos vindo a analisar não pode ser aqui feita de forma que nos satisfaça.

Assim, relativamente ao conjunto de dados dos rebentos de Pau-Brasil remetemos para Diggle (2003) a descrição do ajuste de um modelo de agregação através de medidas resumo e para Baddeley e Turner (2005, 2006) o ajuste de um outro modelo compatível através de máxima pseudoverosimilhança.

Relativamente aos dados do pinheiro negro Japonês a hipótese de aleatoriedade espacial completa foi aceite, indicando-nos o caminho da modelação a seguir.

Quanto aos dados dos fogos vários modelos foram considerados, tendo-se finalmente optado por um modelo de Poisson não homogéneo, que conseguiu capturar bem a tendência observada nos dados.

4.6 Processos pontuais espaço-temporais

Quando os dados a serem analisados são da forma $(x_i, t_i), i = 1, \ldots, n$, onde x_i denota a localização e t_i o correspondente instante de ocorrência de acontecimentos de interesse, e assumimos que os dados formam um registo completo de todos os acontecimentos que ocorreram numa região pré-especificada W e num período de tempo $[0, T]$, dizemos que este conjunto de dados constitui um padrão espaço-temporal e o modelo estocástico que lhe está subjacente designa-se por processo pontual espaço-temporal.

Muitos exemplos da epidemiologia ilustram estes processos, quando estamos interessados, por exemplo, em doenças cuja ocorrência tem uma evolução tanto espacial como temporal.

Várias das ferramentas usadas para analisar os dados de processos espaciais podem ser adaptadas a este contexto espaço-temporal. As medidas resumo baseadas em momentos de pequena ordem podem ser generalizadas de forma óbvia considerando as configurações de acontecimentos nas posições espaciais e temporais especificadas.

No entanto, este cenário espaço-temporal abre caminho para novas formas de modelação e estratégias de análise, levando em consideração de forma mais explicita o caracter unidireccional do tempo - Diggle (2005), Baddeley *et al.* (2006).

Referências

Referências

1. L. Anselin. *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht, 1988.
2. L. Anselin. Local indicators of spatial association - lisa. *Geographical Analysis*, 27:93–115, 1995.
3. B. C. Arnold and D. J. Strauss. Bivariate distributions with conditionals in prescribed exponential families. *Journal of the Royal Statistical Society B*, 53:365–375, 1991.
4. A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, editors. *Case Studies in Spatial Point Process Modeling*. Springer, New York, 2006.
5. A. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics*, 42:283–322, 2000.
6. A. Baddeley and R. Turner. spatstat: an `R` package for analysing spatial point patterns. *Journal of Statistical Software*, 12:1–42, 2005.
7. A. Baddeley and R. Turner. *Case Studies in Spatial Point Pattern Modelling*, chapter Modelling spatial point patterns in R, pages 23–74. Springer, New York, 2006.
8. S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, New-York, 2004.
9. S. Banerjee and A. E. Gelfand. On smoothness properties of spatial processesd. *Journal of Multivariate Analysis*, 84:85–100, 2003.
10. M. S. Bartlett. *Stochastic Processes*. Cambridge University Press, Cambridge, 1955.

11. M. S. Bartlett. *The Statistical Analysis of Spatial Pattern*. Chapman & Hall, London, 1975.
12. M. Berman and T. R. Turner. Approximating point process likelihoods via glim. *Applied statistics*, 41:31–38, 1992.
13. J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B*, 34:192–236, 1974.
14. J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24:179–195, 1975.
15. J. Besag. Efficiency of pseudo-likelihood estimators for simple gaussian fiels. *Biometrika*, 64:616–618, 1977.
16. J. Besag. Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute*, 47:77–92, 1978.
17. J. Besag, R. Milne, and S. Zachary. Point process limits of lattice processes. *Journal of Applied Probability*, 19:210–216, 1982.
18. R. Bivand. More on spatial data. *R News*, 1:13–17, 2001.
19. G. E. P. Box and D. R. Cox. An analysis of transformations (with discussion). *Journal of the Royal Statistics Society, Series B*, 26:211–252, 1964.
20. D. Brook. On the distinction between the conditional probability and the joint probability approaches in the specification of nearest neighbour systems. *Biometrika*, 51:481–483, 1964.
21. A. G. Chetwymd and P. Diggle. On estimating the reduced second moment measure of a stationary spatial point process. *Australian and New Zealand Journal of Statistics*, 40:11–15, 1998.
22. J. M. Clayton, D. G. e Kaldor. Empirical bayes estimates of age standardized relative risks for use in disease mapping. *Biometrics*, 43:671–681, 1987.
23. A. D. Cliff and J. K. Ord. *London Papers in Regional Science 1. Studies in Regional Science*, chapter The problem of spatial autocorrelation, pages 25–55. Pion, London, 1969.
24. A. D. Cliff and J. K. Ord. *Spatial Autocorrelation*. Pion, London, 1973.
25. A. D. Cliff and J. K. Ord. *Spatial Processes: Models and Applications*. Pion, London, 1981.

26. N. Cressie and D. M. Hawkins. Robust estimation of the variogram. *Journal of the International Association for Mathematical Geology*, 12:115–125, 1980.
27. N. A. C. Cressie. *Statistics for Spatial Data*. Wiley, New York, 1991.
28. N. A. C. Cressie. *Statistics for Spatial Data - 2nd edition*. Wiley, New York, 1993.
29. P. Diggle. *Spatial Point Patterns*. Arnold, London, 1983.
30. P. Diggle. A kernel method for smoothing point process data. *Applied Statistics*, 34:138–147, 1985.
31. P. Diggle. *Spatial Point Patterns - 2nd edition*. Arnold, London, 2003.
32. P. Diggle. Spatio-temporal point processes: Methods and applications. Technical Report 78, Johns Hopkins University, 2005.
33. P. Diggle, T. Fiksel, P. Grabarnik, Y. Ogata, D. Stoyan, and M. Tanemura. On parameter estimation for pairwise interaction point processes. *International Statistical Review*, 62:99–117, 1994.
34. P. J. Diggle and P. J. Ribeiro Jr. *Model-Based Geostatistics*. Springer, New York, 2007.
35. J. Durbin. Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables. *Econometrica*, 38:410–421, 1970.
36. L. S. Gandin. Objective analysis of meteorological field. Technical report, Gidrometeorologicheskoe Izdate'stvo., 1963.
37. R. C. Geary. The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5:115–145, 1954.
38. S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.
39. J. W. Gibbs. *Elementary Principles in Statistical Mechanics, Developed with Especial Reference to the Rational Foundation of Thermodynamics*. Charles Scribner's Sons, New York, 1902.
40. D. W. Goodall. Quantitative aspects of plant distribution. *Biological Review*, 27:194–245, 1952.
41. D. W. Goodall. Statistical plant ecolgy. *Anual review of ecology and sistematics*, 1:99–124, 1970.

42. J. M. Hammersley and P. Clifford. Markov fields on finite graphs and lattices. 1971.
43. D. Harte. Statistical seismology library (`SSLib`). 2003.
44. L. Held, I. Natário, S. Fenton, H. Rue, and N. Becker. Towards joint disease mappings. *Statistical Methods in Medical Research*, 14:61–82, 2004.
45. I. Hodder and C. Orton. *Spatial Analysis in Archeology*. Cambridge university press, London, 1976.
46. A. G. Journel and C. J. Huijbregts. *Mining Geostatistics*. Academic Press, London, 1978.
47. J. T. Kent. Continuity properties of random fields. *Annals of Probability*, 17:1432–1440, 1989.
48. P. K. Kitadinis and R. W. Lane. Maximum likelihood parameter estimation of hydrologic spatial processes by the gauss-newton method. *Journal of Hydrology*, 79:53–71, 1985.
49. D. G. Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal Chemical Metallurgical and Mining Society of South Africa*, 52:119–139, 1951.
50. H. W. Lotwick and B. W. Silverman. Methods for analysing spatial processes of several types of points. *Journal of the Royal Statistical Society B*, 44:406–413, 1982.
51. K. V. Mardia and R. J. Marshall. Maximum likelihood estimation of models for residual covariance in apatial regression. *Biometrika*, 71:135–146, 1984.
52. J. Mateu. Obtaining spatial information using spatial point processes. Technical Report Notas de curso, Campus Riu Sec, 2004.
53. G. Matheron. *Traité de Géostatistique Appliquée*. Editions Technip., 1962.
54. G. Matheron. The theory of regionalized variables and its applications. Technical report, Cahiers du Centre de Morphologie Mathematique, Fontainebleu, France, 1971.
55. B. Matérn. Spatial variation. Technical report, Meddelanden fran Statens Skogsforskninginstitut, 1960.
56. B. Matérn. *Spatial Variation*. Springer, New York, 1986.
57. J. Møller and R. P. Waagepetersen. *Statistical Inference and Simulation*

for Spatial Point Processes. Chapman & Hall, Boca Raton, 2003.

58. P. A. P. Moran. The interpretation of statistical maps. *Journal of the Royal Statistical Society B*, 10:243–251, 1948.
59. A. S. Mugglin and B. P. Carlin. Hierarchical modelling in geographic information systems: Population interpolation over incompatible zones. *Journal Agric. Biol. Environ. Statist.*, 3:111–130, 1998.
60. A. S. Mugglin, B. P. Carlin, and A. E. Gelfand. Fully model based approaches for spatially misaligned data. *Journal of the American Statistical Association*, 95:877–887, 2000.
61. I. Natário. Modelos hierárquicos bayesianos para a variação espacial no risco de doença. In Neves and etal., editors, *A Estatística em Movimento, Actas do VIII Congresso Anual da SPE*, pages 261–270, 2001.
62. I. Natário. *Hierarchical Bayesian Models for Epidemiological Analysis of Rare Events*. PhD thesis, Faculdade de Ciências da Universidade de Lisboa, 2005.
63. I. Natário and M. L. Carvalho. Como identificar taxas de doença excessivamente elevadas? In M. L. Carvalho and etal., editors, *Em Novos Rumos da Estatística, Actas do IX Congresso Anual da SPE*, pages 285–296, 2002.
64. I. Natário and M. L. Carvalho. Modelação do cancro de estômago em portugal com recurso a informação indirecta. In Braumann and etal., editors, *Actas do XII Congresso Anual da SPE*, pages 477–488, 2005.
65. I. Natário and L. Knorr-Held. Non-parametric ecological regression and spatial variation. *Biometrical Journal*, 45:670–688, 2003.
66. J. Neyman and E. L. Scott. Statistical approach to problems of cosmology. *Journal of the royal statistical society B*, 20:1–29, 1958.
67. M. Numata. Forest vegetation in the vicinity of choshi. coastal flora and vegetation at choshi, chiba prefecture iv. Technical Report 3, Bulletin of Choshi Marine Laboratory, Chiba university, 1961.
68. Y. Ogata. Statistical modelling for standard seismicity and detection of anomalies by residual analysis. *Tecnophysics*, 169:159–174, 1989.
69. H. D. Patterson and R. Thompson. Recovery of inter-block information when blok sizes are unequal. *Biometrika*, 58:545–554, 1971.
70. H. D. Patterson and R. Thompson. G maximum likelihood estimation

of components of variance. In *Proceedings of the 8th International biometric conference, Biometric Society,*, 1974.

71. R. D. Peng. Multi-dimensional point process models in R. *Journal of Statistical Software*, 8:1–27, 2003.
72. E. C. Pielou. *Mathematical Ecology.* Wiley, New York, 1977.
73. B. D. Ripley. The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13:255–266, 1976.
74. B. D. Ripley. Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society B*, 39:172–212, 1977.
75. B. D. Ripley. *Spatial Statistics.* Wiley, New York, 1981.
76. B. D. Ripley. *Developments in Numerical Ecology*, chapter Spatial point pattern analisys in ecology, pages 407–429. Springer, Berlin, 1987.
77. B. D. Ripley. *Statistical Inference for Spatial Processes.* Cambridge university press, Cambridge, 1988.
78. B. D. Ripley. Spatial statistics in r. *R News*, 1:14–15, 2001.
79. B. D. Ripley and F. P. Kelly. Markov point processes. *Journal of the London Mathematical Society*, 15:188–192, 1977.
80. B. S. Rowlingson and P. Diggle. Splancs: spatial point pattern analysis code in s-plus. *Computers in Geosciences*, 19:627–655, 1993.
81. D. Ruppert, M. P. Wand, and R. J. Carrol. *Semiparametric Regression.* Cambridge University Press, Cambridge, 2003.
82. J. Snow. *On the Mode of Communication of Cholera, 2nd edition.* 8vo, London, 1855.
83. M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging.* Springer, New York, 1999.
84. D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and Its Applications, 2nd edn.* Wiley, New York, 1995.
85. D. Stoyan and H. Stoyan. *Fractals, Random Shapes and Point Fields.* Wiley, New York, 1994.
86. D. J. Strauss. A model for clustering. *Biometrika*, 62:467–475, 1975.
87. G. J. G. Upton and B. Fingleton. *Spatial Data Analysis by Example, Volume 1: Point Pattern and Quantitative Data.* Wiley, Chichester, 1985.
88. M. N. M. Van Lieshout. *Markov Point Processes and Their Applications.*

Imperial college press, London, 2000.

89. P. Whittle. On stationary processes on the plane. *Biometrika*, 41:434–449, 1954.

90. H. Zhang and D. Zimmerman. Toward reconciling two asymptotic frameworks in spatial statistics. *Biometrika*, 92:921–936, 2005.

ISBN 978-972-8890-18-6
Depósito Legal nº 280538/08