

Teste de Aleatoriedade Espacial com Distâncias e Ângulos

Gabriela Drummond Marques da Silva

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE CIÊNCIAS EXATAS
DA
UNIVERSIDADE FEDERAL DE MINAS GERAIS
PARA
OBTENÇÃO DO TÍTULO
DE
MESTRE EM CIÊNCIAS

Área de Concentração: Estatística
Orientador: Prof. Dr. Renato Martins Assunção

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da FAPEMIG

Minas Gerais, fevereiro de 2010

Teste de Aleatoriedade Espacial com Distâncias e Ângulos

Este exemplar corresponde à redação
final da dissertação devidamente corrigida
e defendida por Gabriela Drummond M. da Silva
e aprovada pela Comissão Julgadora.

Banca Examinadora:

- Prof. Dr. Renato Martins Assunção (orientador) - UFMG.
- Prof. Dr. Ilka Afonso Reis - UFMG.
- Prof. Dr. Alexandre Loureiro Rodrigues - UFES.

Agradecimentos

Agradeço primeiramente aos meus pais, por serem mais que pais. Obrigada por serem sempre meus amigos queridos, por me deixarem saber que, por mais independente que eu seja, sempre estarão ali por perto para me amparar.

Às minhas irmãs por serem umas garotas sapecas, por trazerem alegria e amizades incondicionais para a minha vida.

Às minhas avós por serem mulheres corajosas e fortes e por terem me ensinado a importância de correr com garra atrás dos meus sonhos.

Ao meu namorado Renato por ter sido tão atencioso e por ter participado de todos os momentos, me consolando, me ouvindo e cuidando de mim.

A todos os meus amigos e demais familiares por tudo e inclusive pelo sorriso encorajador que compartilharam comigo durante essa caminhada.

Aos professores do Departamento de Estatística da UFMG, em especial, aos professores Rosângela, Glaura, Gregório e Sueli, pelos ensinamentos e pelos exemplos de profissionalismo.

À FAPEMIG- Fundação de Amparo à Pesquisa do Estado de Minas Gerais- pelo auxílio financeiro concedido durante o período dessa dissertação.

Agradeço principalmente ao meu orientador, professor Renato Assunção, pela paciência e por ter me ensinado que a educação é importante mas que criatividade e a curiosidade são as ferramentas que nos levam mais longe. Obrigada pela excelente orientação, pela motivação e pela confiança!

Resumo

A execução de testes de completa aleatoriedade espacial para análise de padrões pontuais localizados em uma região planar é importante em diversas áreas do conhecimento, tais como ecologia e epidemiologia. Inicialmente, os trabalhos que tratavam desse tema no contexto das regiões em que o mapeamento completo é inviável foram desenvolvidos através do chamado Método de Distância. Nessa metodologia, as unidades amostrais são pontos arbitrários selecionados na região em estudo e a inferência estatística é baseada nas distâncias entre cada ponto e os eventos observados mais próximos a ele.

[Assunção \(1994\)](#) e [Assunção e Reis \(2000\)](#) introduziram uma abordagem diferente, o Método de Ângulo. Nesse caso, a informação utilizada é o ângulo formado entre os vetores que ligam o ponto arbitrário aos seus primeiro e segundo eventos mais próximos. Os testes baseados no Método de Ângulo apresentam um poder menor que aqueles classificados no Método de Distância. Este trabalho tem o objetivo de unir as medidas já propostas na literatura, distância e ângulo, em novos testes de hipóteses para aleatoriedade espacial. Dentre os cenários simulados, a proposta resultou em testes mais poderosos do que os testes atualmente vigentes.

Palavras-chave: Métodos de distância, Métodos de Ângulo, aleatoriedade espacial, padrões pontuais.

Abstract

Running complete spatial randomness tests for analysis of point patterns located in a planar region is important in many areas of knowledge, such as ecology and epidemiology. Initially, studies focused in this issue in the context of regions in which the complete mapping is not feasible were developed through the so called Distance Method. In this methodology, sampling units are arbitrary points selected in the area of interest and statistical inference is based on distances between each point and its two nearest events.

Assunção(1994) and Assunção and Reis(2000) introduced a different approach, the Method of Angle. From this perspective, the information used is the angle between the vectors connecting the arbitrary point to its first and second nearest events. The tests based on the Angle Method presented a power smaller than those classified in the Distance Method. This paper aims to join the measures already proposed in the literature, distance and angle, in new tests for spatial randomness. In the simulated scenarios, the proposal resulted in more powerful tests than those currently used.

Keywords: Distance Methods, Angle Methods, spatial randomness, point patterns.

Sumário

Lista de Figuras	vi
1 Introdução	1
1.1 Motivação	1
1.2 Revisão da Literatura	2
1.3 Objetivos	5
1.4 Organização do Trabalho	5
2 Amostragem Esparsa	6
2.1 Método de Distância	7
2.2 Método de Ângulo	9
3 Contribuição	11
3.1 Método de Área	11
3.2 Distribuição da Estatística de Teste	15
4 Simulações	22
5 Conclusões	26
Referências Bibliográficas	28

Lista de Figuras

1.1	Representação dos eventos (P, Q e T, inclusive) e de um ponto de coleta (O) distribuídos em uma determinada região	4
3.1	Interpretação geométrica da estatística de teste H_F	12
3.2	Formas geométricas que compõem as estatísticas de teste do Método de Área	14
3.3	Limites de Integração de W_{j2} para cálculo da marginal de W_{j1}	16
3.4	Distribuição de probabilidade de $2\pi\lambda D_{j1}^2 \cos \theta_j$	18
3.5	Representação dos valores da secante para vários ângulos e da inversa da secante relacionada	21
4.1	Exemplos de realizações do Processo de Thomas.	23
4.2	Poder dos testes sob a hipótese alternativa de padrão agregado.	24

Capítulo 1

Introdução

A seguir será apresentada a motivação desse trabalho por meio da descrição de algumas das áreas do conhecimento em que a modelagem de dados pontuais é importante. Os principais testes de hipótese para completa aleatoriedade espacial serão citados em uma breve revisão da literatura. Finalmente, serão definidos os objetivos do trabalho e as propostas secundárias.

1.1 Motivação

A análise de dados que podem ser representados como pontos no espaço é importante em diversas áreas do conhecimento. Um exemplo é a ecologia vegetal que trata das relações entre esses organismos e o ambiente em que estão inseridos. Um ponto crucial de tal ciência é o estudo da distribuição dos seres vivos utilizando, portanto, técnicas de estatística espacial. Segundo [Li e Zhang \(2007\)](#), a forma como as árvores estão localizadas está relacionada com o crescimento e, por isso, com a produtividade do ecossistema. Essa informação é importante, inclusive, na administração de florestas voltadas para a produção de madeira pois, conhecendo os padrões espaciais que mais favorecem o crescimento é possível definir estratégias para maior rentabilidade do empreendimento. Não apenas para as atividades extrativistas, a análise da localização de plantas auxilia na manutenção e inventário dos biomas ([ErfaniFard et al., 2008](#)). [Shapcott \(1995\)](#) destaca ainda que o acompanhamento das florestas pode auxiliar tanto no desenho de áreas para conservação ambiental quanto nos delineamentos amostrais, úteis na seleção de plantas para uso em programas de melhoria genética.

Como citado em [Diggle \(2003\)](#), o desenvolvimento científico impulsionou o estudo de dados pon-

tuais situadas em regiões planares em áreas tais como a neuroanatomia e a epidemiologia. Na neuroanatomia, são feitas pesquisas que, por exemplo, acompanham o crescimento e o posicionamento de células cancerígenas na seção de um tecido nervoso. A epidemiologia, por sua vez, é a análise geográfica de incidência de doenças (Gatrell *et al.*, 1996) e, usualmente, examina os locais de residência para os casos observados em uma determinada área.

Como citado em Gatrell *et al.* (1996) a análise de processos pontuais apresenta dificuldades na epidemiologia pois ignora a mobilidade dos indivíduos, o que implica a irreal suposição de que uma pessoa não pode contrair doenças em lugares diferentes de sua própria residência. Apesar disso, o estudo de padrões de pontos tem sido empregado na estimação rápida de tamanho de amostras como descrito em Boestoen *et al.* (2007).

Outro exemplo de aplicação é na Criminologia. No Brasil, especificamente, verifica-se o progresso das análises criminológicas como meio de combater a violência no país (Assuncao *et al.*, 2007). Para isso, é interessante compreender a distribuição espacial dos delitos, caracterizados pelos pontos de sua ocorrência nos centros urbanos.

Assim, é visível que a análise espacial de padrões pontuais é necessária em diferentes dimensões da realidade, desde o nível microscópico até o macroscópico. Em todas essas situações, pode-se utilizar testes de hipótese para verificar se o conjunto de dados pontuais se dispõe aleatoriamente em um espaço bidimensional. Esse é o tema da próxima seção que aborda os principais testes de hipótese para aleatoriedade espacial quando apenas dados amostrais podem ser coletados.

1.2 Revisão da Literatura

Como descrito em Diggle (2003), existem inúmeras razões que motivam um pesquisador a executar o teste de completa aleatoriedade espacial logo no início de suas análises. Primeiramente, só faz sentido procurar um modelo para os dados após descartada a possibilidade de que os mesmos estejam distribuídos aleatoriamente. Além disso, esse teste ajuda a explorar as informações disponíveis e propor alternativas à aleatoriedade, operando até mesmo como divisor entre as hipóteses de agregação e regularidade. Por ora, pode-se dizer que a agregação ocorre quando há uma grande concentração de eventos em apenas algumas regiões, enquanto a regularidade é aparente quando os eventos tendem a ter distâncias iguais até os seus vizinhos mais próximos. Posteriormente, os con-

ceitos de aleatoriedade e agregação, os mais importantes para esse trabalho, serão melhor definidos conforme a explicitação dos mecanismos estocásticos geradores desses padrões.

Existem vários trabalhos aplicados à ecologia cujo objetivo principal é verificar a hipótese de que os dados estão distribuídos de forma completamente aleatória espacialmente. Em [ErfaniFard et al. \(2008\)](#) estuda-se uma região da selva iraniana de Zagros para a qual dez testes de hipótese foram feitos e comparados entre si. [Wolf \(2005\)](#) verifica a presença de aleatoriedade espacial para compreender os padrões temporais de mudança em uma floresta localizada na Dinamarca durante um período de cinquenta anos. Nesses exemplos as áreas de interesse foram completamente mapeadas embora em muitas situações isso não seja possível. A despeito do desenvolvimento das técnicas de geoprocessamento, casos em que o interesse está em árvores com características muito específicas, não captadas por imagens de satélite, exigem que o pesquisador colete os dados em campo.

É possível distinguir duas etapas na amostragem de dados pontuais em um determinado território. Nas análises florestais, por exemplo, quando a região é muito grande, a primeira fase define a parcela da mata a ser analisada e a segunda estabelece quais as árvores serão amostradas dentro da parcela escolhida. A última, que corresponde ao escopo desse trabalho, engloba as técnicas aqui chamadas de amostragem esparsa, tema central do Capítulo 2. Usualmente, os elementos do grupo selecionado ao final desse processo são referidos como eventos a fim de distingui-los dos demais pontos arbitrários da área em questão.

Dentre os métodos de amostragem esparsa disponíveis, o foco desse trabalho está na amostragem centrada em pontos. Tal tópico foi introduzido de forma independente por [Skellam \(1952\)](#) e [Moore \(1954\)](#) os quais recomendaram a amostragem aleatória de m localizações da região em estudo, chamadas pontos de coleta. Resumidamente, a inferência estatística dos testes é baseada em dados obtidas no entorno dos pontos de coleta e na compatibilidade dos valores observados e esperados sob a hipótese de aleatoriedade. O benefício desse procedimento de amostragem é o sorteio de pontos aleatórios, o que elimina a necessidade de enumeração da população de eventos.

As técnicas de amostragem mais tradicionais dentre as centradas em pontos, chamadas Método de Distância, propõem o uso da distância dos pontos de coleta até seus eventos vizinhos como medida de informação. Na figura 1.1, O é um ponto de coleta e P e Q são o primeiro e o segundo

eventos mais próximos de O. Os segmentos OP e OQ são exemplos das quantidades utilizadas, as distâncias do ponto amostral até o evento mais próximo e até o segundo evento mais próximo. Nas próximas seções serão descritos com mais detalhes alguns testes de hipótese pertencentes à classe dos Métodos de Distância, como os sugeridos por [Besag e Gleaves \(1973\)](#) e [Hines e Hines \(1979\)](#), os mais proeminentes da literatura em termos de poder. Especialistas em botânica se apropriaram muito bem dessas técnicas. [Li e Zhang \(2007\)](#), por exemplo, descreveram a disposição de abetos vermelhos do noroeste dos Estados Unidos valendo-se da proposta de [Besag e Gleaves \(1973\)](#). Atualmente, é possível visualizar aplicações também na estimativa do tamanho de populações humanas em situações de emergência ([Grais *et al.*, 2006](#) e [Boestoen *et al.*, 2007](#)).

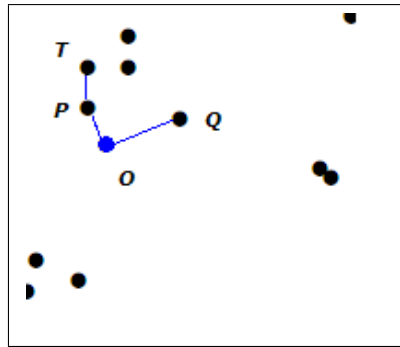


Figura 1.1: Representação dos eventos (P, Q e T, inclusive) e de um ponto de coleta (O) distribuídos em uma determinada região

A utilização dos Métodos de Distância para amostragem das áreas estudadas não sofreu modificações até a década de 90, quando [Assunção \(1994\)](#) introduziu uma nova concepção do problema, por meio do uso do ângulo θ formado por POQ. Sob a hipótese de aleatoriedade, θ é uniformemente distribuído em $(0, \pi)$. Esse teste será melhor explicado nas próximas seções juntamente com algumas de suas modificações propostas em [Assunção e Reis \(2000\)](#) e [Staupendahl e Zucchini \(2006\)](#).

Diante da tecnologia disponível no período em que foi proposto, o ângulo θ possuía vantagens de aplicação em campo. Quando os eventos mais próximos ao ponto de coleta estavam visíveis, tornava-se mais fácil calcular θ que as demais medidas e, em outras circunstâncias, o esforço na obtenção de θ era apenas marginal. Já existem trenas eletrônicas que simplificam cálculo das distâncias nos mais diversos cenários mas, devido ao custo, elas ainda não se encontram amplamente disponíveis. Dessa forma, o ângulo ainda é atual e seu uso pode ser visto em [Graz \(2006\)](#) e [Crecente-Campo *et*](#)

al. (2009).

1.3 Objetivos

Este trabalho pretende unir algumas das propostas já feitas pela literatura sugerindo novos testes de hipótese para completa aleatoriedade espacial que utilizem simultaneamente a medida do ângulo e das distâncias entre o ponto de coleta e o primeiro e segundo eventos mais próximos a ele. Os objetivos específicos são:

- revisar os principais testes de hipótese para completa aleatoriedade espacial que consideram a amostragem centrada em pontos,
- propor um novo teste de hipótese nessa classe,
- encontrar a distribuição da estatística de teste proposta,
- comparar os métodos da literatura e o sugerido nesse trabalho.

1.4 Organização do Trabalho

Esta dissertação está disposta da seguinte forma: no Capítulo 2, define-se amostragem esparsa e, mais especificamente, os Métodos de Distância e de Ângulo. No Capítulo 3 são apresentados os teste de hipótese que correspondem à contribuição desse trabalho para o tema e no Capítulo 4 o poder dos teste propostos são comparados com o poder daqueles já existentes na literatura. Finalmente, no Capítulo 5 faz-se um resumo dos objetivos pretendidos e discute-se os principais resultados alcançados.

Capítulo 2

Amostragem Esparsa

Na amostragem tradicional seleciona-se uma única sub-área a ser completamente mapeada pois os resultados obtidos para ela representam também as conclusões esperadas para a região em que está inserida. O mapeamento completo das áreas, entretanto, pode requerer uma soma grande de recursos humanos e financeiros. Como opção ao mapeamento completo, surgiu então a amostragem esparsa. Por meio dela, coleta-se dados em localizações dispersas por toda a região de abrangência e não apenas uma grande massa de informação em uma pequena localidade.

Segundo [Diggle \(2003\)](#) existem dois tipos de amostragem esparsa: a centrada em pontos, definida no Capítulo 1, e a via *quadrats*. Os *quadrats* são polígonos de mesmo tamanho e forma, alocados aleatoriamente na área em questão, que delimitam pequenas parcelas do território a serem completamente mapeadas. O procedimento que faz uso de tais polígonos é preferível, em relação à amostragem centrada em pontos, para a detecção de padrões agregados nos quais os aglomerados são grandes, formando um caso particular de heterogeneidade ([Grabarnik e Chiu, 2007](#)). A utilização dos *quadrats* em campo, entretanto, requer atenção, já que resultados obtidos podem diferir de acordo com o extensão das parcelas escolhida pelo pesquisador ([de Souza Lima-Ribeiro, 2007](#)). Essa é uma decisão difícil pois implica em considerar a relação entre a forma e o tamanho do *quadrat* com a escala da distribuição de eventos. Quando os eventos são esparsos, por exemplo, as parcelas exigidas seriam relativamente grandes e, por isso, de complicada delimitação em um ambiente florestal ([Liu, 2001](#)).

Uma vez que a amostragem centrada em pontos não pretende substituir o mapeamento completo dos dados, a facilidade e rapidez com que as informações são obtidas a tornam atrativa em um

estágio de análise preliminar, principalmente em grandes populações. Embora seja muito útil em alguns casos, a coleta centrada em pontos possui algumas desvantagens. Como citado em [Liu \(2001\)](#), as informações tendem a ser processadas apenas para os vizinhos imediatamente mais próximos. Desse forma, apenas as nuances do processo nas menores escalas são captadas, e as escalas maiores são ignoradas. Além disso, o poder da maior parte dos métodos baseados em pontos de coleta é muito baixo quando a hipótese alternativa é que o padrão dos dados é regular. Logo, não existe um método amostral uniformemente melhor. A escolha irá depender dos recursos financeiros e tempo disponíveis assim como da configuração esperada para dados *a priori*.

2.1 Método de Distância

A hipótese de completa aleatoriedade espacial é equivalente a supor que o mecanismo estocástico gerador do padrão de pontos é o Processo de Poisson Homogêneo (PPH). Considere \mathcal{A} uma região em estudo e λ a intensidade, ou número médio de eventos por unidade de área. Sejam ainda m o número de pontos de coletas sorteados em \mathcal{A} e D_{ji} a distância entre o j -ésimo ponto de coleta, $j = 1, \dots, m$, e o i -ésimo evento mais próximo a esse, $i = 1, 2$. Segundo as propriedades de PPH e ignorando efeitos de borda, a distribuição da variável aleatória πD_{j1}^2 é Exponencial (λ). Essa informação, entretanto, é insuficiente para originar um teste de hipótese, pois a distribuição encontrada depende de λ que, em geral, é um parâmetro desconhecido. Tal problema foi solucionado por [Holgate \(1965\)](#) ao propor um teste baseado na média amostral da razão $Z_{sti} = d_{j1}^2/d_{j2}^2$, da seguinte forma:

$$H_N = m^{-1} \sum (d_{j1}^2/d_{j2}^2).$$

Essa estatística de teste, aqui chamada de H_N , é distribuída aproximadamente de acordo com uma Normal com média 1 e variância $12m^{-1}$ sob a hipótese de aleatoriedade. Logo, não depende de quaisquer parâmetros desconhecidos e pode ser adotada facilmente. Uma variante da estatística de [Holgate \(1965\)](#) é:

$$H_F = \sum d_{j1}^2 / \sum (d_{j2}^2 - d_{j1}^2)$$

que possui aproximadamente distribuição $F_{2m, 2m}$.

A idéia de compor estatísticas independentes de λ por meio da comparação de duas medidas

de distância diferentes foi explorada em outros trabalhos. Um artigo de destaque foi o de [Besag e Gleaves \(1973\)](#), que propôs o teste T^2 . Nele os autores utilizam, de forma original, a distância entre eventos em uma área de busca restrita, uma vez que, até então, apenas a distância entre os pontos de coleta e os seus eventos mais próximos havia sido considerada. Este teste é ainda mais poderoso que aquele sugerido por [Holgate \(1965\)](#) e, para compreendê-lo melhor, observe a Figura 1.1. Seja T o evento vizinho mais próximo de P , sob a restrição de que ângulo OPT possui pelo menos 90° . O teste T^2 se baseia nas distâncias OP (d_{j1}) e PT (t_{j1}). As duas possíveis estatística de teste são:

$$T_N = m^{-1} \sum d_{j1}^2 / (d_{j1}^2 + t_{j1} / 2)$$

e

$$T_F = 2 \sum d_{j1}^2 / \sum t_{j1}^2$$

as quais apresentam distribuições sob o PPH iguais as suas estatísticas de Holgate correspondentes, H_N e H_F .

[Hines e Hines \(1979\)](#) sugeriu uma bem sucedida modificação da estatística T^2 , que é da seguinte forma:

$$H_I = 2m \frac{\sum (2d_{j1} + t_{j1})}{\sum (\sqrt{2d_{j1}} + \sqrt{t_{j1}})^2}.$$

A estatística de [Hines e Hines \(1979\)](#) possui um poder semelhante aos testes de [Besag e Gleaves \(1973\)](#) sob a hipótese alternativa de regularidade, mas um poder superior a este sob a hipótese alternativa de agregação. Os valores críticos de H_I são encontrados também em [Hines e Hines \(1979\)](#).

Ainda existe o interesse no aprimoramento desses testes. Uma mostra disso é o trabalho de [Liu \(2001\)](#), que modifica estatísticas de testes baseadas em distâncias já conhecidas, observando seu comportamento para j de ordem maior que 2. Assim, o poder dos cinco testes modificados é comparado via simulações Monte Carlo para os padrões alternativos regular, agregado e heterôgeneo. O padrão heterôgeneo corresponde ao Processo de Poisson não Homogêneo, que não será abordado nesse trabalho. Além disso, [Boestoen et al. \(2007\)](#) propõem um aprimoramento do teste T^2 durante o processo de seleção de eventos. As informações obtidas entretanto, não são utilizadas para fins

de testes de hipótese e sim para estimação do número de eventos na população.

2.2 Método de Ângulo

Uma nova abordagem para testes de hipóteses em regiões que não podem ser completamente mapeadas foi proposta por [Assunção \(1994\)](#). Sabe-se que o ângulo θ formado por POQ na Figura 1.1 possui distribuição $U(0, \pi)$ sob a hipótese nula. Entretanto, quando um padrão de agregação ocorre, o ângulo observado tende a ser menor que o esperado sob o PPH. Para uma compreensão intuitiva, considere um padrão cujos eventos possuem uma agregação bastante alta. De acordo com a seleção aleatória dos pontos de coleta, uma vez que a região fora dos conglomerados tenderá a ser maior, há maior probabilidade de que eles sejam alocados nesse local. Logo, os primeiros e segundos eventos mais próximos possivelmente farão parte da borda do conglomerado mais próximo e, por isso, terão uma distância pequena entre si.

Supondo que os ângulos $\theta_1, \dots, \theta_m$ de diferentes fontes amostrais são independentes, o teste proposto para avaliar a aleatoriedade foi o teste de Kolmogorov. Durante as simulações, entretanto, o teste do ângulo apresentou poder menor que T_N , possivelmente por se basear em um teste global que não está associado a alternativas específicas. Outro problema é que como o número de pontos de coleta deve ser restrito a 10% do total de eventos para garantia da independência entre os ângulos amostrais ([Byth e Ripley, 1980](#)), o número de unidades amostrais tende a ser pequeno e os resultados assintóticos pouco confiáveis.

Em [Assunção e Reis \(2000\)](#), procura-se solucionar esse problema obtendo-se mais informações de um mesmo ponto de coleta e também utilizando outros testes de aderência para a distribuição uniforme. O aumento de dados é feito pela inclusão dos ângulos com os terceiros e quartos eventos mais próximos ao ponto de coleta. Esse teste do ângulo modificado obteve poder similar ao teste de [Besag e Gleaves \(1973\)](#).

Os testes baseados em ângulos tiveram uma boa aceitação por parte dos ecólogos que o utilizaram e modificaram. Em [Triković e Yamamoto \(2008\)](#), por exemplo, os autores propõem a média de $\theta_1, \dots, \theta_m$ como índice de dispersão. Para isso, sugerem a comparação da média amostral com o seu valor esperado sob a hipótese de aleatoriedade.

Outra idéia muito utilizada pelos ecólogos foi a substituição do valor absoluto pelo uso de uma medida de referência calculada por simulação Monte Carlo (Staupendahl e Zucchini, 2006). O índice proposto então é a fração dos ângulos na amostra menores que o ângulo padrão. A distribuição de frequência dessa estatística de teste é calculada e o teste Qui-quadrado é utilizado para observar a aderência dos valores observados à distribuição esperada. A vantagem desse método é que, na maior parte dos casos, é possível concluir sem quaisquer ferramentas, apenas pela observação, se um ângulo é maior ou menor que o limite crítico determinado.

Já em Corral-Rivas *et al.* (2010) o ângulo é utilizado na estatística de teste com um formato bem diferente. Inicialmente, a amostragem centrada em pontos surge de uma maneira não usual em que o ponto de coleta não é a unidade amostral e sim o evento mais próximo a ele. A partir daí, o índice denominado média direcional é obtido para cada elemento da amostra. Para isso calcula-se a soma R_j de vetores unitários cujo ponto de origem é a árvore referência e cuja direção é orientada para um dos seus quatro vizinhos mais próximos. Seja $\alpha_{j2}, \dots, \alpha_{jn}$ o valor dos ângulos entre o primeiro e cada um dos demais n vizinhos mais próximos, respectivamente e no sentido horário, é possível mostrar que R_j é da forma:

$$R_j = \sqrt{\left(1 + \sum_{i=2}^n \cos \alpha_{ji}\right)^2 + \left(\sum_{i=2}^n \sin \alpha_{ji}\right)^2}$$

A estatística final é a média dos R_j que possui o valor crítico encontrado por meio de simulações, assim como em Staupendahl e Zucchini (2006).

Capítulo 3

Contribuição

A seguir serão apresentados os testes de hipótese para a completa aleatoriedade espacial sugeridos nesse trabalho. A interpretação geométrica dessas propostas será discutida e a distribuição assintótica da estatística de teste principal será provada.

3.1 Método de Área

Uma vez que a proposta desse trabalho se alicerçará em argumentos geométricos, antes de mais nada, é importante estabelecer uma nova visão sobre os testes já definidas na literatura. Sabe-se, por exemplo, que a estatística H_F é expressa por:

$$H_F = \sum_j d_{j1}^2 / \sum_j (d_{j2}^2 - d_{j1}^2).$$

A informação pertencente à equação anterior também pode ser obtida de forma visual recorrendo-se à Figura 3.1. Nela, os pequenos círculos sólidos em preto são os dois eventos mais próximos ao ponto de coleta, representado, por sua vez, pelo quadrado situado no centro da imagem. Considerando ainda, o círculo C_{j1} de raio D_{j1} e o círculo C_{j2} com raio D_{j2} , então H_F pode ser interpretada como a comparação entre as áreas do círculo C_{j1} e do anel hachurado $AN_j = C_{j2} \cap C_{j1}$, $j = 1, \dots, m$, da seguinte forma:

$$\begin{aligned} H_F &= \frac{\sum_j \text{área de } C_{j1}}{\sum_j \text{área de } AN_j} \\ &= \frac{\sum_j \pi d_{j1}^2}{\sum_j \pi (d_{j2}^2 - d_{j1}^2)} = \frac{\sum_j d_{j1}^2}{\sum_j (d_{j2}^2 - d_{j1}^2)}. \end{aligned}$$

Quanto maior a diferença entre o tamanho dessas áreas, maior será o indício de um padrão agregado. Isso ocorre uma vez que, quando um padrão agregado é observado, os eventos mais próximos a ele tendem a ter uma distância pequena entre si, como visto no Capítulo 2. Soma-se a isso o fato de que, com alta probabilidade, o ponto de coleta estará situado fora do conglomerado em que os seus eventos mais próximos estão localizados. Assim, D_{j1} tenderá a ser grande em relação a $D_{j2} - D_{j1}$, o que leva à alta diferença entre a área do círculo C_{j1} e do anel AN_j

A estatística H_F pode ser útil para identificar configurações como na Figura 3.1(a), mas é pouco eficiente quando há um padrão como na Figura 3.1(b). Note que a diferença entre os cenários não pode ser captada apenas pela distância do ponto de coleta até os eventos, mesmo que essa seja uma informação importante. Embora em ambos os casos as distâncias sejam as mesmas, na última situação os dois eventos não parecem pertencer a um mesmo conglomerado e, por isso, há maior propensão de que venham de um PPH.

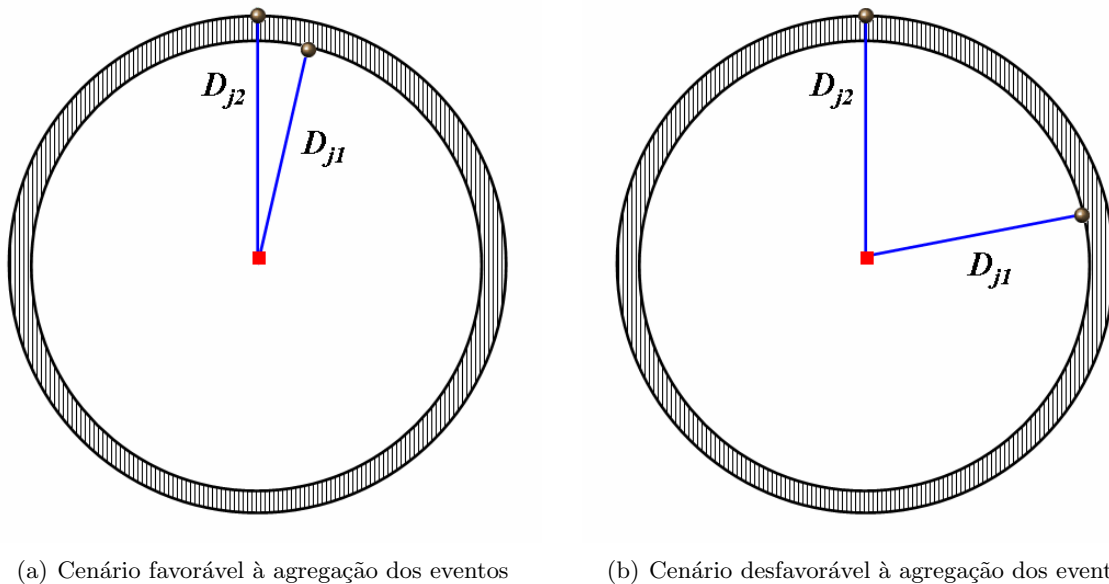


Figura 3.1: Interpretação geométrica da estatística de teste H_F

Para solucionar esse problema Besag e Gleaves (1973) incluiu a distância entre eventos nas estatísticas de teste e, para que houvesse independência dessas medidas com relação a D_{j1} e D_{j2} , utilizou a área de busca restrita. O ângulo sugerido por Assunção (1994), entretanto, é capaz de distinguir as diferenças entre as Figuras 3.1(a) e 3.1(b) sem a necessidade da coleta de distâncias.

Em campo, isso permite que o pesquisador permaneça em sua base de observação inicial se os dois eventos mais próximos estão visíveis. Não é necessário ainda utilizar a área de restrição de T^2 pois θ_j é independente de D_{j1} e D_{j2} , sob a hipótese nula.

Assim como o teste T^2 não emprega somente a distância entre os eventos, o interessante é que o ângulo seja visto como informação complementar à D_{j1} e D_{j2} e não uma fonte de dados isolada. Dessa forma surgiu o Método de Área que é a proposta desse trabalho e que une as medidas de distância e ângulo nos testes de hipótese para completa aleatoriedade espacial. Isso é feito por meio de estatísticas de teste que correspondem à razão da área de diferentes formas geométricas compostas pelas medidas D_{j1} , D_{j2} e θ_j . São elas:

$$A_1 = \frac{\sum (d_{j1} * \cos \theta_j)^2}{\sum (d_{j2}^2 - d_{j1}^2)},$$

$$A_2 = \frac{\sum (d_{j1}^2 * \cos \theta_j)}{\sum (d_{j2}^2 - d_{j1}^2)}.$$

A interpretação dessas estatísticas pode ser vista com auxílio da Figura 3.2. Nela, assim como anteriormente, o quadrado no centro representa um ponto de coleta e os pequenos círculos sólidos são os primeiro e segundo eventos mais próximos a ele. Além disso, as duas imagens possuem a mesma escala e, por isso, utilizando-se as indicações de distância da Figura 3.2(a), pode-se distinguir, na Figura 3.2(b), os círculos C_{j1} e C_{j2} , de raios D_{j1} e D_{j2} , e o anel hachurado AN_j , todos eles já definidos na Figura 3.1. Acrescenta-se a esses, entretanto, a presença do círculo C_j cujo raio $D_{j1} \cos \theta_j$ é a projeção de D_{j1} em D_{j2} .

A área de C_j é $\pi(D_{j1} * \cos \theta_j)^2$ enquanto a área de AN_j é $\pi(D_{j2}^2 - D_{j1}^2)$. Logo, a estatística de teste A_1 corresponde à razão entre a soma das áreas de C_j e a soma das áreas de AN_j , para todos os pontos de coleta, da seguinte forma:

$$A_1 = \frac{\sum \text{área de } C_j}{\sum \text{área de } AN_j}$$

$$= \frac{\sum \pi (d_{j1} \cos \theta_j)^2}{\sum \pi (d_{j2}^2 - d_{j1}^2)} = \frac{\sum (d_{j1} \cos \theta_j)^2}{\sum (d_{j2}^2 - d_{j1}^2)}.$$

Por meio da Figura 3.2(b), é possível observar que a área do retângulo R_j , de base D_{j1} e altura

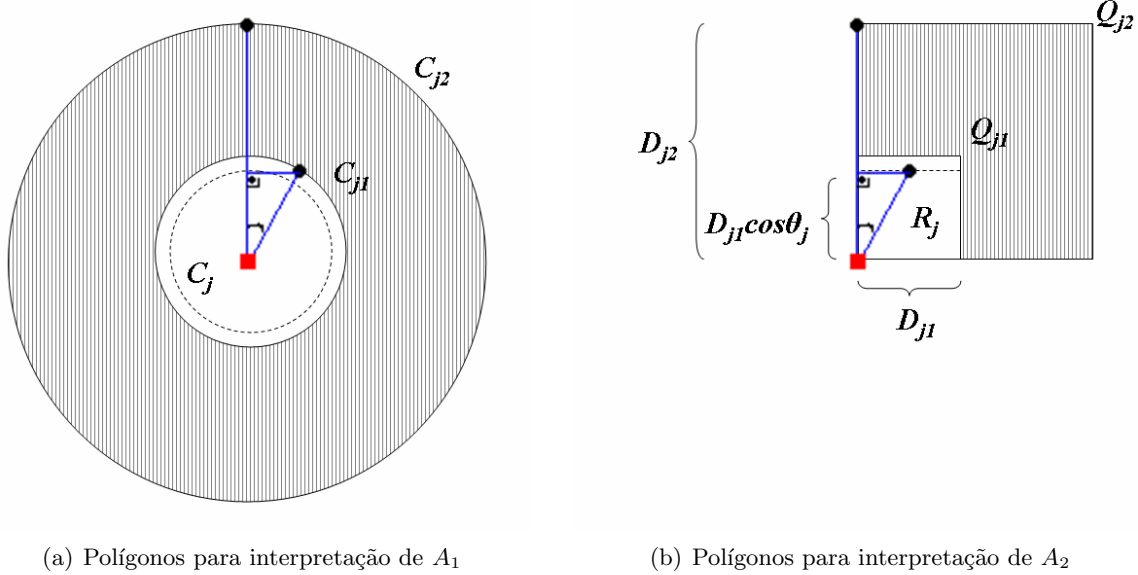
(a) Polígonos para interpretação de A_1 (b) Polígonos para interpretação de A_2

Figura 3.2: Formas geométricas que compõem as estatísticas de teste do Método de Área

$D_{j1} \cos \theta$, é $D_{j1}^2 \cos \theta$. Na mesma figura, estão representados ainda os quadrados Q_{j1} e Q_{j2} de lados D_{j1} e D_{j2} , respectivamente. A estatística A_2 é a razão da soma das áreas de R_j e a soma das área de Q_j para todos os pontos de coleta, sendo Q_j a diferença das áreas de Q_{j2} e Q_{j1} :

$$\begin{aligned} A_2 &= \frac{\sum \text{área de } R_j}{\sum (\text{área de } Q_{j2} - \text{área de } Q_{j1})} \\ &= \frac{\sum (d_{j1}^2 \cos \theta_j)}{\sum (d_{j2}^2 - d_{j1}^2)}. \end{aligned}$$

Para entender o funcionamento dessas estatísticas, basta pensar em A_1 e A_2 como uma versão ponderada da tradicional estatística H_F , tendo $\cos \theta_j$ como fator de ponderação. Intuitivamente, pode-se imaginar que, quando um padrão agregado é observado, os primeiros e segundos eventos mais próximos ao ponto de coleta estarão localizados no mesmo conglomerado e D_{j1} será aproximadamente igual a D_{j2} , como visto na Figura 3.1(a). Por isto, a diferença $D_{j2}^2 - D_{j1}^2$ tenderá a ser pequena e H_F apresentará um valor maior que o esperado sob a hipótese de aleatoriedade. Esse fenômeno, entretanto, não é o único observado quando os dois eventos mais próximos ao ponto de coleta vêm do mesmo adensamento. Como discutido no Capítulo 2, em padrões agregados, o ângulo θ_j tenderá a ser pequeno e, por isso, $\cos \theta_j$ próximo de 1. Assim, quanto mais forte a agregação, maior também serão as estatísticas A_1 e A_2 . Logo, A_1 e A_2 utilizam mais informações para con-

cluír sobre a natureza dos eventos. Se, por exemplo, um ponto de coleta tiver os seus primeiros e segundos eventos à uma distância parecida, mas em direções muito diferentes, como visto na Figura 3.1(b), a evidência de agregação é menor do que a esperada para configurações como na Figura 3.1(a). Tal evidência é captada corretamente por A_1 e A_2 mas não por H_F .

3.2 Distribuição da Estatística de Teste

Dentre as estatísticas de testes propostas, aquela que possui os melhores resultados empíricos é A_2 . Esse resultado está na Seção 4, onde o poder assintótico dessa estatística, que se mostrou aproximadamente igual ao poder empírico, é apresentado. Logo, é importante obter a sua distribuição de probabilidade calculando-se, primeiramente, a distribuição de $\cos \theta_j$, que compõe A_2 . Sabe-se que, sob a hipótese nula, $\theta_j \sim U(0, \pi)$. Como $X = \cos \theta_j$ é uma função monótona decrescente no intervalo $[0, \pi]$:

$$F_X(x) = P(\cos \theta_j \leq x) = P(\theta_j > \cos^{-1} x) = 1 - P(\theta_j \leq \cos^{-1} x) = 1 - \frac{\cos^{-1} x}{\pi}.$$

Assim, para obter a densidade de X basta derivar a equação acima:

$$f_X(x) = \frac{dF_X(x)}{dx} = \frac{1}{\pi\sqrt{1-x^2}}; -1 < x < 1.$$

Para encontrar a distribuição de $D_{j1}^2 \cos \theta_j$ pode-se utilizar ainda $Y_{j1} = 2\pi\lambda D_{j1}^2$ que é independente de θ_j sob a hipótese nula e, segundo Diggle (2003), segue uma Qui-quadrado com 2 graus de liberdade. Seja ainda $Y_{j2} = \cos \theta_j$, a distribuição de $W_{j1} = Y_{j1}Y_{j2}$ pode ser calculada via método do Jacobiano, tendo como variável auxiliar $W_{j2} = Y_{j2}$ pois ambas são funções bijetores e resultam e um Jacobiano cujas derivadas parciais existem. Nesse caso, a distribuição conjunta de W_{j1} e W_{j2} é:

$$\begin{aligned} f_{W_{j1}, W_{j2}}(w_{j1}, w_{j2}) &= f_{Y_{j1}, Y_{j2}}\left(w_{j2}, \frac{w_{j1}}{w_{j2}}\right) \left| \frac{1}{J(y_{j1}, y_{j2})} \right|, \text{ se } w_{j2} > 0 \text{ e } 1 < \frac{w_{j1}}{w_{j2}} < 1 \\ &= \frac{e^{-w_{j2}/2}}{2w_{j2}\pi\sqrt{1-(w_{j1}/w_{j2})^2}} = \frac{e^{-w_{j2}/2}}{2\pi\sqrt{w_{j1}^2 - w_{j2}^2}}. \end{aligned}$$

A fim de que a distribuição de W_{j1} seja encontrada corretamente, deve-se avaliar os limites de integração de W_{j2} , indicados pela área destacada na Figura 3.3 e obtidos considerando-se as seguintes condições de existência:

$$\begin{aligned} w_{j1} &: -\infty \leq w_{j1} < \infty \\ w_{j2} &: w_{j2} \geq 0 \\ \frac{w_{j1}}{w_{j2}} &: -1 < \frac{w_{j1}}{w_{j2}} < 1 \Rightarrow -w_{j2} < w_{j1} < w_{j2}. \end{aligned}$$

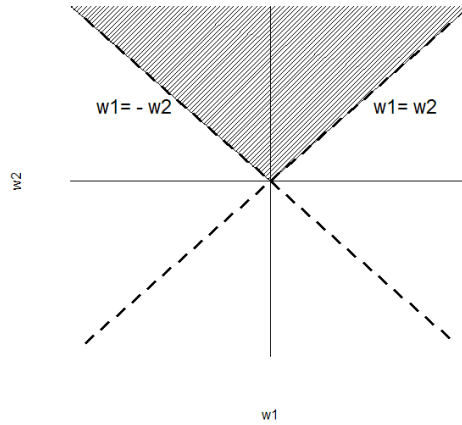


Figura 3.3: Limites de Integração de W_{j2} para cálculo da marginal de W_{j1}

Logo, a distribuição marginal de interesse é:

$$f_{W_{j1}}(w_{j1}) = \begin{cases} \int_{-w_{j1}^+}^{\infty} \frac{e^{-w_{j2}/2}}{2\pi\sqrt{w_{j1}^2 - w_{j2}^2}} dw_{j2} & \text{se } w_{j1} \leq 0 \\ \int_{w_{j1}^+}^{\infty} \frac{e^{-w_{j2}/2}}{2\pi\sqrt{w_{j1}^2 - w_{j2}^2}} dw_{j2} & \text{se } w_{j1} > 0 \end{cases}$$

Substituindo-se $x = w_{j1}/w_{j2}$, o que implica que $w_{j2} = xw_{j1}$ e $dw_{j2} = dxw_{j1}$, os novos limites

de integração são:

$$w_{j1} \leq 0 : \begin{cases} \text{se } w_{j2} = +\infty \text{ então } x = -\infty \\ \text{se } w_{j2} = -w_{j1} \text{ então } x = -1 \\ \text{como } w_{j2} \geq 0 \text{ se } w_{j1} \leq 0 \text{ então } x \leq 0. \end{cases} ,$$

$$w_{j1} > 0 : \begin{cases} \text{se } w_{j2} = +\infty \text{ então } x = +\infty \\ \text{se } w_{j2} = w_{j1} \text{ então } x = 1 \\ \text{como } w_{j2} \geq 0 \text{ se } w_{j1} > 0 \text{ então } x > 0. \end{cases} .$$

Assim, a distribuição de W_{j1} pode ser reescrita como:

$$f_{W_{j1}}(w_{j1}) = \begin{cases} \int_{-1^-}^{-\infty} w_{j1} \frac{e^{-xw_{j1}/2}}{2\pi\sqrt{x^2w_{j1}^2 - w_{j1}^2}} dx \text{ se } w_{j1} \leq 0 \\ \int_{1^+}^{\infty} w_{j1} \frac{e^{-xw_{j1}/2}}{2\pi\sqrt{x^2w_{j1}^2 - w_{j1}^2}} dx \text{ se } w_{j1} > 0 \end{cases} .$$

$$f_{W_{j1}}(w_{j1}) = \begin{cases} \int_{-1^-}^{-\infty} \frac{e^{-xw_{j1}/2}}{2\pi\sqrt{x^2 - 1}} dx \text{ se } w_{j1} \leq 0 \\ \int_{1^+}^{\infty} \frac{e^{-xw_{j1}/2}}{2\pi\sqrt{x^2 - 1}} dx \text{ se } w_{j1} > 0 \end{cases} .$$

A relação entre os limites de integração e as condições de existência dessas duas equações, permite que elas sejam representadas na forma única:

$$f_{W_{j1}}(w_{j1}) = \int_{1^+}^{\infty} \frac{e^{-x|w_{j1}|/2}}{2\pi\sqrt{x^2 - 1}} dx, \quad -\infty \leq w_{j1} < \infty \quad (3.1)$$

Esse resultado remete à função de Bessel modificada do segundo tipo de ordem ν que, segundo [Abramowitz e Stegun \(1972\)](#), possui a seguinte representação via integral:

$$K_\nu(z) = \frac{\pi^{1/2}(1/2z)^\nu}{\Gamma(\nu + 1/2)} \int_1^\infty e^{-zt}(t^2 - 1)^{\nu-1/2} dt, \text{ se } \mathcal{R}\nu > -1/2$$

sendo $\mathcal{R}\nu$ a parte real de ν , se ν é complexo. Considerando $\nu = 0$:

$$\begin{aligned} K_0(z) &= \frac{\pi^{1/2}}{\Gamma(1/2)} \int_1^\infty e^{-zt} (t^2 - 1)^{-1/2} dt \\ &= \int_1^\infty \frac{e^{-zt}}{\sqrt{(t^2 - 1)}} dt. \end{aligned}$$

Finalmente, é possível aplicar a função de Bessel diretamente na Equação 3.1, obtendo-se então:

$$f_{W_{j1}}(w_{j1}) = \frac{K_0(|w_{j1}|/2)}{2\pi}, \quad -\infty \leq w_{j1} < \infty.$$

A função de Bessel possui diversas representações, sendo que a mais usual delas é na forma:

$$K_0(z) = -\ln\left(\frac{z}{2}\right) I_0(z) + \frac{1}{2} \sum_{k=0}^{\infty} \left\{ 2 \times \psi(k+1) \frac{\left(\frac{1}{4}z^2\right)^k}{2k!} \right\}$$

onde ψ é a função Digamma e $I_0(z)$ é a função de Bessel modificada do primeiro tipo, ambas definidas em Abramowitz e Stegun (1972). Trata-se portanto, de uma soma infinita de séries de potências que, por isso, só pode ser delineada graficamente via integração numérica. Utilizando esse recurso, a Figura 3.4 representa a distribuição marginal de $W_{j1} = 2\pi\lambda D_{j1}^2 \cos\theta_j$.

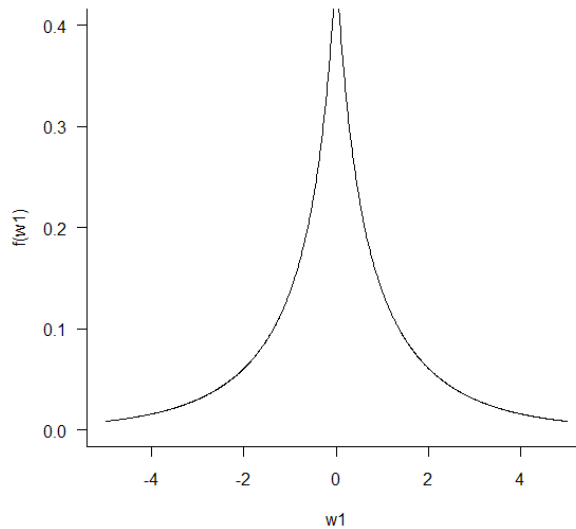


Figura 3.4: Distribuição de probabilidade de $2\pi\lambda D_{j1}^2 \cos\theta_j$

É possível observar que a função de densidade de $W_{j1} = 2\pi\lambda D_{j1}^2 \cos \theta_j$ é simétrica e centrada em zero. Além disso, dentre as funções $k_\nu(x)$, aquela com ν igual a zero apresenta a maior inclinação.

Diante dos resultados obtidos, resta ainda encontrar a distribuição da soma $S_{1m} = \sum_{j=1}^m W_{1j} = 2\pi\lambda \sum D_{j1}^2 * \cos \theta_j$, uma vez que a distribuição $2\pi\lambda \sum_j (D_{j2}^2 - D_{j1}^2)$ é conhecida na literatura. Pelo Teorema Central do Limite, sejam $W_{j1}, j = 1 \dots m$, variáveis aleatórias independentes e identicamente distribuídas, com média 0 e variância σ^2 , é possível dizer que:

$$\sqrt{m} \frac{S_{1m}}{m} \xrightarrow{d} N(0, \sigma^2), \text{ quando } m \rightarrow \infty.$$

Uma vez que a $E(W_{j1}) = 0$:

$$\sigma^2 = Var(w_{j1}) = E(w_{j1}^2) = \int_{-\infty}^{\infty} w_{j1}^2 \left[\int_{1+}^{\infty} \frac{e^{-x|w_{j1}|/2}}{2\pi\sqrt{x^2-1}} dx \right] dw_{j1}.$$

Utilizando o Teorema de Fubini pode-se inverter a ordem de integração. Logo:

$$\begin{aligned} \sigma^2 &= \int_{1+}^{\infty} \frac{1}{2\pi\sqrt{x^2-1}} \left[\int_{-\infty}^{\infty} w_{j1}^2 e^{-x|w_{j1}|/2} dw_{j1} \right] dx \\ &= \int_{1+}^{\infty} \frac{1}{2\pi\sqrt{x^2-1}} \left[\int_{-\infty}^0 w_{j1}^2 e^{-x(-w_{j1}/2)} dw_{j1} + \int_0^{\infty} w_{j1}^2 e^{-xw_{j1}/2} dw_{j1} \right] dx \\ &= \int_{1+}^{\infty} \frac{1}{2\pi\sqrt{x^2-1}} \left[2 \int_0^{\infty} w_{j1}^2 e^{-xw_{j1}/2} dw_{j1} \right] dx. \\ &= \int_{1+}^{\infty} \frac{2}{2\pi\sqrt{x^2-1}} \left(\frac{2}{x} \right) \left[\int_0^{\infty} \left(\frac{x}{2} \right) w_{j1}^2 e^{-w_{j1}x/2} dw_{j1} \right] dx. \end{aligned}$$

A integral em W_{j1} corresponde à $E(Y^2)$ se Y segue uma distribuição Exponencial com média $\lambda = x/2$. Sabe-se que, nesse caso,

$$E(Y^2) = \frac{2}{\lambda^2} = \frac{8}{x^2}.$$

Portanto, a variância de W_{j1} é:

$$\begin{aligned}\sigma^2 &= \int_{1+}^{\infty} \frac{2}{\pi x \sqrt{x^2 - 1}} \left(\frac{8}{x^2} \right) dx. \\ &= \frac{8}{\pi} \int_{1+}^{\infty} \frac{1}{x^3 \sqrt{x^2 - 1}} dx.\end{aligned}$$

Segundo Spiegel (1968), a forma indefinida da integral obtida anteriormente corresponde a:

$$\int \frac{1}{x^3 \sqrt{x^2 - 1}} dx = \frac{\sqrt{x^2 - 1}}{2x^2} + \frac{1}{2} \sec^{-1} |x|.$$

Aplicando os limites de integração, tem-se que:

$$\begin{aligned}\sigma^2 &= \frac{8}{\pi} \left[\left(\lim_{x \rightarrow \infty} \frac{\sqrt{x^2 - 1}}{2x^2} + \lim_{x \rightarrow \infty} \frac{1}{2} \sec^{-1} |x| \right) \right. \\ &\quad \left. - \left(\lim_{x \rightarrow 1+} \frac{\sqrt{x^2 - 1}}{2x^2} + \lim_{x \rightarrow 1+} \frac{1}{2} \sec^{-1} |x| \right) \right].\end{aligned}$$

Para solucionar essa equação é necessário aplicar a regra de L'Hopital em:

$$\lim_{x \rightarrow \infty} \frac{\sqrt{x^2 - 1}}{2x^2} = \lim_{x \rightarrow \infty} \frac{1}{8x \sqrt{x^2 - 1}} = 0.$$

Além disso, precisa-se definir claramente a função inversa da secante, que é:

$$y = \sec^{-1}(z) \Leftrightarrow z = \sec(y), \quad y \in \left[0, \frac{\pi}{2}\right) \cup \left(\frac{\pi}{2}, \pi\right].$$

Os valores de y para cada z estão representados na Figura 3.5. Diante dos resultados anteriores, torna-se possível calcular a variância de W_{j1} ,

$$\sigma^2 = \frac{8}{\pi} \left(0 + \frac{1}{2} \frac{\pi}{2} \right) (0 + 0) = 2.$$

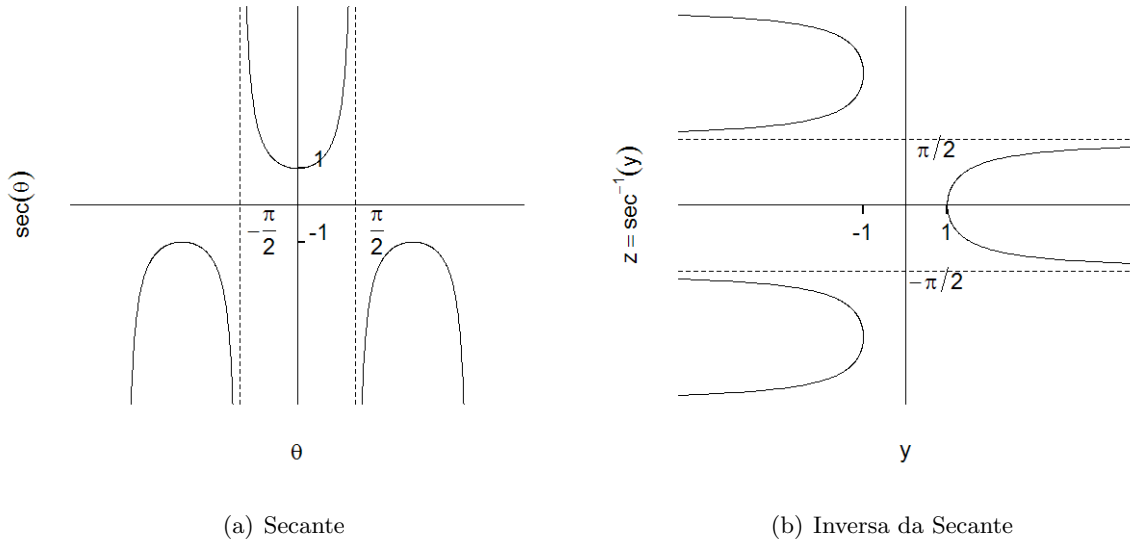


Figura 3.5: Representação dos valores da secante para vários ângulos e da inversa da secante relacionada

Assim, pode-se concluir que a distribuição assintótica de $S_{1m} = 2\pi\lambda \sum D_{j1}^2 * \cos \theta_j$ é:

$$\sqrt{m} \frac{S_{1m}}{m} \xrightarrow{d} N(0, 2), \text{ a medida que } m \rightarrow \infty. \quad (3.2)$$

Sabe-se ainda, segundo [Holgate \(1965\)](#) que os $Z_j = 2\pi\lambda(D_{j2}^2 - D_{j1}^2), j \dots m$, são variáveis aleatórias independentes cuja distribuição é Qui-Quadrado com 2 graus de liberdade. Logo, de acordo com a Lei Forte dos Grandes Números, pode-se dizer que, com probabilidade 1,

$$\frac{S_{2m}}{m} = \frac{2\pi\lambda \sum_j (D_{j2}^2 - D_{j1}^2)}{m} \xrightarrow{q.c.} 2, \text{ quando } m \rightarrow \infty. \quad (3.3)$$

Finalmente, utilizando as Equações 3.2 e 3.3 é possível encontrar a distribuição de A_2 :

$$\sqrt{m} A_2 = \sqrt{m} \frac{S_{1m}/m}{S_{2m}/m} \xrightarrow{d} \frac{1}{2} N(0, 2), \text{ com } m \rightarrow \infty,$$

ou seja,

$$\sqrt{m} A_2 \xrightarrow{d} N\left(0, \frac{2}{\sqrt{2}}\right), \text{ a medida que } m \rightarrow \infty.$$

Capítulo 4

Simulações

Com o objetivo de avaliar a qualidade dos testes de hipótese propostos nesse trabalho, o poder de cada um deles foi calculado. A hipótese alternativa utilizada durante os cálculos foi a de agregação, uma vez que a regularidade quase não é observada na natureza e, por isso, não é de grande interesse nesse estudo. Processos seguindo a hipótese de agregação podem ser gerados construtivamente como um conjunto de pontos originários do chamado Processo de Thomas Modificado (Diggle *et al.*, 1976). Esse processo se inicia com a geração aleatória dos eventos pais em uma região A de acordo com um PPH com intensidade λ . Na segunda etapa, cada evento pai produz independentemente um número aleatório de eventos filhos segundo uma distribuição de Poisson com média μ . Os eventos filhos são alocados independentemente em torno dos seus respectivos pais seguindo uma distribuição normal bivariada, composta por variáveis aleatórias independentes com média 0 e desvio padrão σ .

Exemplos dos cenários utilizados nesse trabalho podem ser vistos na Figura 4.1. É importante lembrar, entretanto, que as imagens representam a configuração completa de eventos embora os métodos abordados abranjam apenas o entorno dos pontos de coleta como região de análise. Durante as simulações, considerou-se $\lambda=0,1$, μ igual 4,6,8 e 10 e σ entre 0,5 e 1, com incrementos de 0,1. Os processos pontuais foram gerados na área $A=[0, 50] \times [0, 50]$, enquanto os pontos de coleta foram sorteados na janela $S=[0, 40] \times [0, 40]$, a fim de evitar possíveis efeitos de borda. Os parâmetros escolhidos permitiram a avaliação da qualidade dos testes propostos tendo como hipóteses alternativas desde padrões com aglomeração muito evidente (Figura 4.1(a)), até padrões em que a agregação dos pontos é mais fraca (Figura 4.1(c)).

Foram utilizadas 5000 repetições Monte Carlo para achar a distribuição empírica da estatística

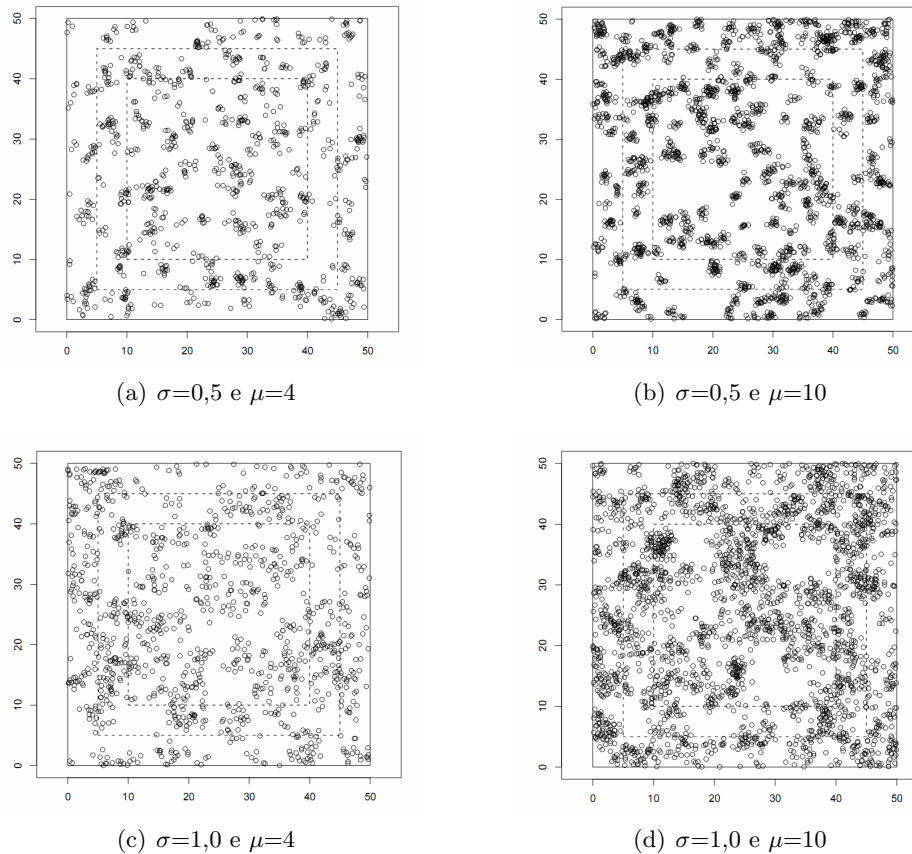


Figura 4.1: Exemplos de realizações do Processo de Thomas.

A_1 tanto sob o PPH quanto sob o Processo de Thomas. Os percentis 0,025 e 0,975 da distribuição encontrada foram estabelecidos como limites críticos para o cálculo do poder. Já para A_2 , o poder foi obtido tomando-se como base a distribuição assintótica de A_2 sob H_0 e utilizando-se uma distribuição empírica sob a hipótese alternativa de agregação.

Os resultados estão descritos na Figura 4.2. Dentre os testes do Método de Distância, o que apresentou melhor desempenho foi H_I . Por esse motivo, os resultados para as estatísticas T_N , T_F , H_N e H_F foram suprimidos das imagens, embora tenham sido calculados. Além disso, a estatística de Assunção (1994) foi utilizada como representante dos testes baseados em ângulo pois vale-se apenas dos dois vizinhos mais próximos ao ponto de coleta e, por isso, tem a região de abrangência similar à de A_1 e A_2 . Se estatísticas que fazem uso de um número de vizinhos maior do que 2 fossem consideradas, seria necessário analisar o esforço na coleta da amostra, que não é o objetivo atual desse trabalho.

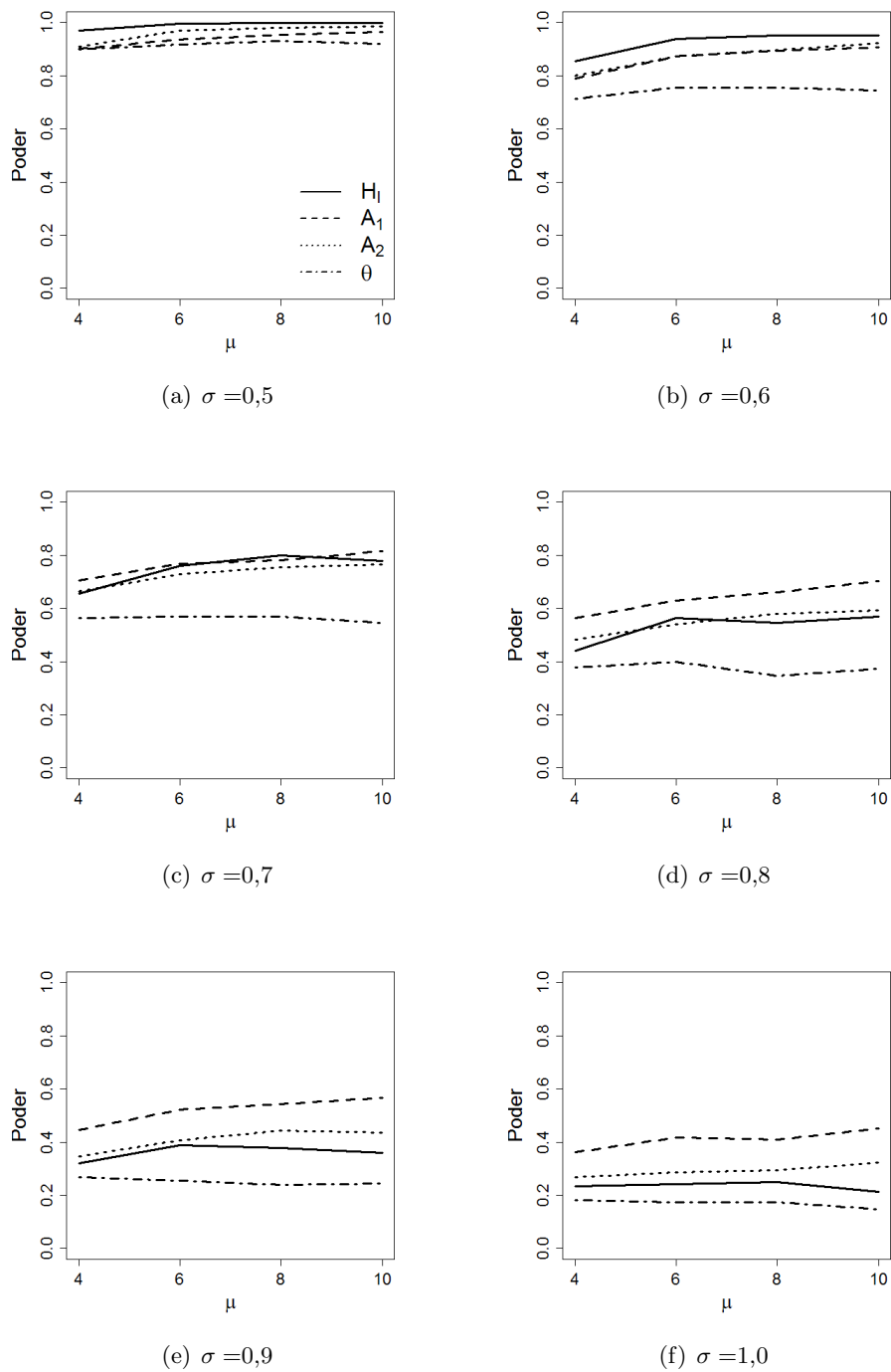


Figura 4.2: Poder dos testes sob a hipótese alternativa de padrão agregado.

Por meio dos gráficos nas Figuras 4.2(a) a 4.2(f) é possível observar que o poder dos testes não sofre grande impacto com a mudança de μ mas é muito influenciado por oscilações no parâmetro σ . Isso se deve ao fato de que, quanto maior o σ , maior a probabilidade que os pontos de coleta

sejam alocados no interior dos conglomerados. Assim, os eventos mais próximos irão se dispor de forma semelhante a um PPH e, por isso, a identificação do padrão de eventos real será dificultada. Este é um exemplo do problema já mencionado em que apenas as pequenas escalas dos processos são objeto de estudo da amostragem centrada em pontos.

Dentre os testes propostos anteriormente pela literatura, H_I é o que possui o poder mais elevado quando σ apresenta os menores valores. À medida que σ cresce, entretanto, todos os testes sofrem uma substancial queda no nível de poder. Por outro lado, para as estatísticas A_1 e A_2 é verificado um poder razoável quando $\sigma < 0,7$ e, além disso, não é possível observar uma queda tão acentuada no poder de acordo com σ . Com isso, A_1 e A_2 destacam-se das outras estatísticas abordadas quando $\sigma > 0,7$.

Os resultados encontrados, indicam A_1 e A_2 , principalmente A_2 , como as melhores opções de estatísticas para o teste de aleatoriedade espacial quando não é possível mapear toda a área de interesse. Isso é verdade, pois eles possuem bons resultados quando σ pode ser considerado pequeno e não apresentam poder tão baixo quanto os demais testes quando σ é grande.

Capítulo 5

Conclusões

Esse estudo teve o objetivo de propor um novo teste de hipótese para completa aleatoriedade espacial na presença de dados pontuais situados em uma região planar. Considerou-se ainda que as informações provêm de uma amostragem esparsa cujas unidades amostrais são pontos de coletas selecionados aleatoriamente na região em análise. Nesse contexto, os testes de hipótese até então presentes na literatura podem ser categorizados nos chamados Métodos de Distância e Métodos de Ângulo. Essa nomenclatura varia de acordo com o tipo de informação requerida, coletada no entorno dos pontos amostrais, seja distância ou ângulo, respectivamente.

As inovações desse trabalho tiveram início com a interpretação geométrica da estatística H_F desenvolvida por [Holgate \(1965\)](#). Por meio disso, foi possível perceber que os testes baseados no Método de Distância não funcionam bem quando os eventos vizinhos ao ponto de coleta possuem distâncias parecidas mas em direções diferentes. A partir daí, identificou-se formas geométricas que seriam úteis nas situações em que esses testes falham. Observou-se, assim, que as estruturas encontradas possuíam simultaneamente o ângulo e as distâncias como componentes de formação.

Os testes sugeridos nesse trabalho utilizam então: 1) a distância D_{ji} do j -ésimo ponto de coleta, $j = 1, \dots, m$, até seu i -ésimo evento vizinho, $i = 1, 2$ e 2) o ângulo θ_j formado entre os primeiros e segundos eventos mais próximos ao ponto de coleta j , tendo o próprio ponto de coleta como origem. A inovação dos testes propostos, entretanto, não foi as medidas em que se basearam e, sim, o fato de utilizá-las conjuntamente. Duas estatísticas de testes, A_1 e A_2 , foram então originadas e iniciaram uma nova classe de análise, o Método de Área, um contraponto aos Método de Distância e Método de Ângulo. O método de Área possui esse nome pois corresponde à razão entre áreas de

formas geométricas constituídas por D_{ji} e θ_j .

Um considerável ganho no poder do teste de completa aleatoriedade espacial foi obtido por meio das estatísticas A_1 e A_2 . Quando os conglomerados do Processo de Thomas, utilizado como hipótese alternativa ao PPH, são estreitos ($\sigma < 0,7$), a maior parte dos testes de hipótese são bons. As estatísticas do Método de Área também não deixam a desejar e registram um poder apenas ligeiramente menor que a estatística de [Hines e Hines \(1979\)](#). Já quando os conglomerados são mais extensos ($\sigma > 0,7$), os testes de hipótese tendem a apresentar um poder pequeno. Nesse caso, A_1 e, principalmente A_2 , se destacam como os testes que possuem os maiores poderes observados.

Em trabalhos futuros seria interessante incluir os dados de mais eventos vizinhos aos pontos de coleta, e não apenas as informações sobre os primeiro e segundo eventos mais próximos. Esse procedimento pode ser útil para minimizar o problema da má detecção do padrão de pontos em escalas maiores. Além disso, seria importante verificar o número de pontos de coleta a partir do qual os resultados assintóticos são confiáveis.

Referências Bibliográficas

- Abramowitz, M. e Stegun, I. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Dover Publications, Nova York. 17, 18
- Assunção, R. M. (1994). Testing spatial randomness by means of angles, *Biometrics* **50**(2): 531–537. iii, 4, 9, 12, 23
- Assunção, R. M. e Reis, I. A. (2000). Testing spatial randomness: a comparison between t2 methods and modifications of the angle test, *Brazilian Journal of Probability and Statistics* **14**(1): 71–86. iii, 4, 9
- Assuncao, R., Tavares, A., Correa, T. e Kulldorff, M. (2007). Space-time cluster identification in point processes, *Canadian Journal of Statistics - La Revue Canadienne de Statistique* **35**: 9–25. 2
- Besag, J. e Gleaves, J. T. (1973). On the detection of spatial pattern in plant communities, *Bulletin of the International Statistical Institute* **45**(1): 153–158. 4, 8, 9, 12
- Boestoen, K., Chalabi, Z. e Grais, R. (2007). Optimisation of the T-square sampling method to estimate population sizes, *Emerging Themes in Epidemiology*. 2, 4, 8
- Byth, K. e Ripley, B. (1980). On sampling spatial patterns by distance methods, *Biometrics* **36**(2): 279–284. 9
- Corral-Rivas, J. J., Wehenkel, C., Castellanos-Bocaz, H. A., Vargas-Larreta, B. e Diéguez-Aranda, U. (2010). A permutation test of spatial randomness: application to nearest neighbour indices in forest stands, *Journal of Forest Research* **15**: 218–225. 10
- Crecente-Campo, F., Pommerening, A. e Rodriguez-Soalleiro, R. (2009). Impacts of thinning on structure, growth and risk of crown fire in a *Pinus sylvestris* L. plantation in northern Spain, *Forest Ecology and Management* **257**(9): 1945–1954. 4
- de Souza Lima-Ribeiro, M. (2007). Distribuição espacial de espécies arbóreas em fragmentos de Cerrado sentido restrito no Planalto Central brasileiro - GO, *Biociências* **15**(2): 160–165. 6
- Diggle, P. J. (2003). *Statistical analysis of spatial point patterns*, 2 edn, Academic Press London, Londres. 1, 2, 6, 15
- Diggle, P. J., Besag, J. e Gleaves, J. T. (1976). Statistical analysis of spatial point patterns by means of distance methods, *Biometrics* **32**(3): 659–667. 22
- ErfaniFard, Y., Fegghi, J., Zobeiri, M. e Namiranian, M. (2008). Comparison of two distance methods for forest spatial pattern analysis (Case Study: Zagros Forest of Iran), *Journal of Applied Sciences* **8**(1): 152–157. 1, 3

- Gatrell, A. C., Bailey, T. C., Diggle, P. J. e Rowlingson, B. S. (1996). Spatial point pattern analysis and its application in Geographical Epidemiology, *Transactions of the Institute of British Geographers* **21**(1): 256–274. [2](#)
- Grabarnik, P. e Chiu, S. N. (2007). Goodness-of-fit test for complete spatial randomness against mixtures of regular and clustered spatial point process, *Biometrika* **89**(2): 411–421. [6](#)
- Grais, R. F., Coulombier, D., Ampuero, J., Lucas, M. E., Barretto, A. T., Jacquier, G., Diaz, F., Balandine, S., Mahoudeau, C. e Brown, V. (2006). Are rapid population estimates accurate? A field trial of two different assessment methods., *Disasters* **30**(3): 364–376. [4](#)
- Graz, F. P. (2006). Spatial diversity of dry savanna woodlands - Assessing the spatial diversity of a dry savanna woodland stand in northern Namibia using neighbourhood-based measures, *Biodiversity and Conservation* **15**(4): 1143–1157. [4](#)
- Hines, W. G. S. e Hines, R. J. O. (1979). The Eberhardt index and the detection of non-randomness of spatial point distributions, *Biometrika* **66**: 73–80. [4](#), [8](#), [27](#)
- Holgate, P. (1965). Some new tests of randomness, *Journal of Ecology* **53**(2): 261–266. [7](#), [8](#), [21](#), [26](#)
- Li, F. e Zhang, L. (2007). Comparison of point pattern analysis methods for classifying the spatial distributions of spruce-fir stands in the north-east USA, *Forestry* **80**(3): 337–349. [1](#), [4](#)
- Liu, C. (2001). A comparison of five distance-based methods for spatial pattern analysis, *Journal of Vegetation Science* **12**: 411–416. [6](#), [7](#), [8](#)
- Moore, P. G. (1954). Spacing in plant populations, *Ecology* pp. 222–227. [3](#)
- Shapcott, A. (1995). The spatial genetic structure in natural populations of the Australian temperate rainforest tree *Atherosperma moschatum* (Labill.) (Monimiaceae), *Heredity* **74**: 28–38. [1](#)
- Skellam, J. G. (1952). Studies in statistical ecology: Spatial pattern, *Biometrika* **39**(3-4): 346. [3](#)
- Spiegel, M. (1968). *Mathematical Handbook of formulas and tables*, McGraw-Hill, Nova York. [20](#)
- Staupendahl, K. e Zucchini, W. (2006). Estimating the spatial distribution in forest stands by counting small angles between nearest neighbors., *Allgemeine Forst und Jagdzeitung* **177**: 160 – 168. [4](#), [10](#)
- Triković, S. e Yamamoto, H. (2008). Indexing of spatial patterns of trees using a mean of angles, *Journal of Forest Research* **13**: 117–121. [9](#)
- Wolf, A. (2005). Fifty year record of change in tree spatial patterns within a mixed deciduous forest, *Forest Ecology and Management* **215**(1-3): 212–223. [3](#)