

Estatística de varredura unidimensional para detecção de conglomerados de acidentes de trânsito em arruamentos

Marcelo Azevedo Costa¹, Marcos Oliveira Prates², Marcos Antônio da Cunha Santos²

¹Departamento de Engenharia de Produção – Universidade Federal de Minas Gerais (UFMG)
Av. Presidente Antônio Carlos, 6627, Cep 30161-010, Belo Horizonte – MG – Brazil

²Departamento de Estatística – Universidade Federal de Minas Gerais (UFMG)

macosta.est@gmail.com, {marcosop,msantos}@est.ufmg.br

***Abstract.** This paper presents a new approach for cluster detection of spatial point patterns which are restricted to street networks. The proposed method is an extension of the temporal scan statistic which is applied to spatial line segments. Geographical coordinates of points are initially mapped into a one dimension geographical structure, which is the geo-coded line of the street of interest. In this dimension, the events are identified by their relative distances to a point of origin. A one-dimensional varying scanning window identifies portions of the street where the incidence rate of car accidents is higher than the expected. Statistical inference is obtained using Monte Carlo simulations. The methodology was implemented in the R software and provides a friendly graphical user interfaces. The software provides online interface with Google maps.*

***Resumo.** Este artigo apresenta uma nova abordagem para a varredura de eventos pontuais espaciais restritos a estruturas de arruamentos. O método proposto é uma extensão do modelo geo-estatístico de varredura temporal mas, considera eventos pontuais espalhados ao longo de um arruamento. Dessa forma, coordenadas geográficas de eventos pontuais são inicialmente mapeadas em uma única dimensão, que é a linha georeferenciada do arruamento de interesse. Nesta dimensão, os eventos pontuais são identificados pelas suas distâncias relativas a um ponto de origem. Uma janela unidimensional e de dimensão variável realiza a varredura no arruamento, procurando identificar trechos nos quais a taxa de incidência de acidentes de trânsito é maior que a esperada. Inferência estatística é obtida a partir de simulações de Monte Carlo. A metodologia foi implementada no software R e utiliza interfaces gráficas e mapas de arruamento obtidos a partir de interfaces com o ambiente Google maps.*

1. Introdução

A estatística de varredura espacial, proposta por Kulldorff (1997), permite a identificação de conglomerados espaciais a partir de eventos pontuais ou eventos de áreas. Dessa forma, a metodologia permite delinear regiões no espaço onde a intensidade da ocorrência de um evento é maior, ou menor que o esperado. Esta metodologia se tornou muito popular, em diversas áreas do conhecimento, como demonstra Costa e Kulldorff (2009). Como consequência, novas abordagens tem sido propostas, como extensões para detecção de conglomerados puramente temporais ou espaço-temporal [Kulldorff *et al.* 1998; Kulldorff, 2001; Kulldorff *et al.*, 2005]. Além de novas metodologias que exploram variações na geometria espacial e espaço-temporal da janela de varredura [Alm, 1997; Kulldorff, 2006; Duczmal and Assunção, 2004; Costa *et al.*, 2012].

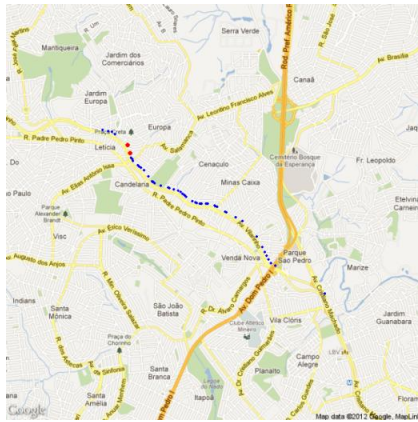
Em particular, este trabalho apresenta uma nova variação da estatística de varredura desenvolvida, a princípio, para a análise de eventos pontuais cuja ocorrência é restrita a estruturas de arruamentos. Análises de conglomerados puramente espaciais aplicados a dados de trânsito podem ser encontrados na literatura [Huang *et al.*, 2009]. Entretanto, uma análise puramente espacial não permite identificar localmente, isto é, ao longo de um arruamento específico, regiões de alta ou maior incidência de eventos pontuais. Por um lado, um cluster puramente espacial poderá abranger diversos arruamentos, sem que haja qualquer diferenciação com relação à contribuição dos eventos de cada arruamento. Como consequência, um trecho de um arruamento poderá ser caracterizado como crítico simplesmente porque a análise de conglomerado não faz distinção quanto a esta característica. É o caso, por exemplo, de um arruamento paralelo à uma avenida que apresenta alta incidência de eventos pontuais. Em particular, a caracterização de trechos críticos de ruas e avenidas permitirá aos órgãos responsáveis a criação de políticas de restrição como radares e melhorias de sinalização.

A metodologia apresentada foi desenvolvida a partir um projeto de pesquisa envolvendo o Centro de Estudos de Criminalidade e Segurança da UFMG (CRISP) e a Empresa de Transporte e Trânsito de Belo Horizonte (BHTRANS). Utilizando dados georeferenciados provenientes de acidentes de trânsito ocorridos no período de 2004 a 2011, foi desenvolvida uma plataforma para consulta, visualização e análises de dados em ambiente R. A plataforma, denominada **RBHTrans** possibilita ao usuário a consulta total ou parcial da base de dados e, a partir dos dados selecionados, disponibiliza funcionalidades de análise de mapas de *kernel*, moda espacial, análise descritiva de eventos de arruamentos e a estatística de varredura linear, denominada *street scan*. A plataforma utiliza os pacotes *RgoogleMaps* e *Rgooglevis* que possibilitam o acesso online a mapas da plataforma Google maps, além da possibilidade de exportar atributos georeferenciados para visualização em ambiente browser, como o Google Chrome ou Mozilla Firefox. Utilizando esta plataforma, o usuário pode realizar análises de arruamentos e visualizar os dados georeferenciados de acidentes sobrepostos a mapas de arruamento, satélite, ou mesmo visualizações utilizando o ambiente *street view* do Google maps.

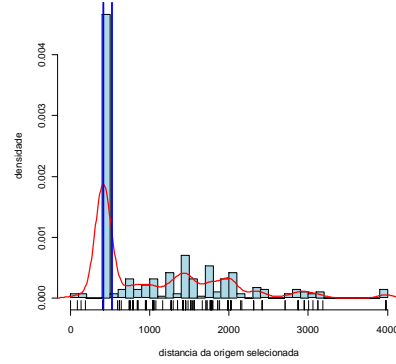
2. A Estatística de Varredura Unidimensional

Seja s_i um par de coordenadas espaciais, $s_i = (x_i, y_i)$, atribuídas a um i -ésimo evento pontual espacial. Seja também $i \in \{1, \dots, N\}$, onde N é o número total de eventos pontuais em um determinado arruamento. Como exemplo, a Figura 1(a) mostra as

coordenadas relativas a acidentes de trânsito ocorridos ao longo da Avenida Vilarinho em Belo Horizonte/MG, no ano de 2010.



(a) Eventos pontuais ao longo de um arruamento. Em vermelho estão indicados os eventos pertencentes a um conglomerado de alta incidência.



(b) Histograma da frequência de eventos pontuais com relação ao primeiro ponto s_1 do arruamento. As linhas verticais delimitam um conglomerado detectado pelo método de detecção de conglomerados em arruamentos.

Figura 1. Eventos pontuais localizados ao longo da Avenida Vilarinho, Belo Horizonte, MG.

Para delimitar o arruamento de interesse, definem-se os eventos s_1 e s_N como o primeiro evento e o último evento, respectivamente, no arruamento de interesse. Dessa forma, estamos interessados em detectar trechos entre os pontos s_1 e s_N que apresentam altas ou baixas intensidades de eventos pontuais. Para isso, devemos definir a distância entre o evento inicial s_1 e o i -ésimo evento s_i no arruamento. Defini-se então $d_{i,i+1}^*$ como a distância euclidiana entre dois eventos consecutivos s_i e s_{i+1} , tal que não exista nenhum outro ponto entre estes dois eventos. Dessa forma, definimos a distância entre o i -ésimo evento e o evento inicial (s_1) como:

$$d_{1,i} \approx \sum_{j=1}^{i-1} d_{j,j+1}^*. \quad (1)$$

Essa aproximação é adequada em situações onde a distância entre pontos consecutivos ao longo de arruamentos seja muito pequena. A Figura 1(b) mostra a distribuição de eventos pontuais ao longo de um arruamento considerando as distâncias relativas de cada evento ao ponto de origem, s_1 .

Seja agora, definida a hipótese nula de que os $N - 2$ eventos localizados entre os pontos s_1 e s_N ocorram de forma homogênea no trecho. Sob esta hipótese, a intensidade de eventos, λ_{H_0} , ao longo do trecho do arruamento é definida por:

$$\lambda_{H_0} = \frac{N-2}{d_{1,N}}. \quad (2)$$

Dessa forma, o número de casos ocorridos ao longo de um trecho de comprimento Δd ($\Delta d < d_{1,N}$) pode ser modelado por variável de Poisson, $Y_{\Delta d} \sim Poisson(\mu_{\Delta d} = \lambda_{H_0} \cdot \Delta d)$. É de particular interesse identificar automaticamente trechos ao longo do arruamento que apresentem um intensidade maior que a esperada. Para isso, propomos o seguinte teste de razão de verossimilhança: seja l uma janela de comprimento variável, tal que $0 < d_l < d_{1,N}$, c_l é o número observado de eventos ao longo de l e μ_l é o número esperado de casos ao longo de l . Sob a hipótese nula temos que $\mu_l = d_l \cdot \lambda_{H_0}$. A estatística do teste da razão de verossimilhança é obtida na forma:

$$\hat{\kappa} = \kappa(\hat{l}) = \sup_l \left(\frac{c_l}{\mu_l} \right)^{c_l} \left(\frac{N - c_l - 2}{N - \mu_l - 2} \right)^{N - c_l - 2} \quad (3)$$

A partir da Equação (3), é possível identificar o trecho \hat{l} que apresenta a maior ou menor incidência de eventos. Caso seja de interesse identificar somente trechos de alta incidência, então deve ser aplicada a restrição: $c_l > \mu_l$. Para avaliar o valor observado da estatística de teste em relação à Hipótese nula, é proposta uma simulação de Monte Carlo:

1. S simulações independentes são realizadas. Para cada simulação $N - 2$ eventos pontuais são homogeneamente distribuídos ao longo de $d_{1,N}$.
2. Para cada simulação a estatística da razão de verossimilhança é calculada, $\hat{\kappa}_1, \dots, \hat{\kappa}_S$.
3. Caso o valor observado da estatística de teste esteja acima do valor do percentil $100(1 - \alpha)\%$ dos valores simulados, então rejeita-se a hipótese nula.
4. Caso a hipótese nula seja rejeitada, pode-se dizer que o trecho \hat{l} detectado é crítico.

3. Implementação Computacional

A estatística de varredura unidimensional foi implementada no software R, e utiliza os pacotes *RgoogleMaps* e *googleVis*. O pacote *RgoogleMaps* [Loecher, 2010] possibilita a importação de imagens do ambiente *Google maps* para o software R. As imagens são importadas no formato png (*Portable Network Graphics*) e são utilizadas como plano de fundo onde é possível a sobreposição de pontos, linhas e polígonos. A importação de mapas e sobreposição da imagem é obtida a partir da seguinte sequência de comandos:

```
R> MyMap <- GetMap.bbox(lonR, latR, center, size = c(640, 640),
destfile = "MyTile.png", ...)
R> PlotOnStaticMap(MyMap, lat, lon, destfile, ...)
```

onde *lonR* e *latR* são os limites de longitude e latitude do mapa a ser obtido, *center* é o parâmetro de centralidade do mapa (opcional), *size* é a resolução da imagem e *destfile* é o nome do arquivo de destino da imagem. No comando *PlotOnStaticMap*, *lat* e *lon* são os vetores de pontos a serem sobrepostos na imagem *MyMap*.

O pacote *googleVis* [Gesmann and de Castillo, 2011] possibilita a exportação de dados em HTML utilizando recursos do Google Visualisation API. Utilizando a funcionalidade *gvisMap()* é possível visualizar dados pontuais utilizando diretamente a plataforma *Google maps*, a partir de um browser, como ilustrado na Figura 2. Neste ambiente, a funcionalidade *street view* do *Google maps* pode ser utilizada para visualizar os dados ao longo do arruamento.



Figura 2. Visualização de dados pontuais no ambiente Google maps, utilizando a funcionalidade `gvisMap()` do pacote *googleVis*. Utilizando o ambiente *street view* é possível visualizar as coordenadas de conglomerados de acidentes de trânsito ao longo do arruamento de interesse.

A metodologia de varredura unidimensional foi implementada na funcionalidade `street_scan()`. O procedimento de simulação de Monte Carlo, que apresenta grande custo computacional, foi implementado em linguagem C e incorporada ao ambiente R na forma de uma *dll* (*Dynamic-link library*) denominada *varredura.dll*. Foram criadas interfaces gráficas para a seleção de atributos do banco de dados bem como a seleção de parâmetros para as funcionalidades: (a) análise da intensidade de eventos em arruamentos, (b) mapa de kernel, (c) moda espacial, (d) análise de séries temporais, (e) *street scan* e (f) visualização e dados. A base de dados, as funcionalidades implementadas, a *dll* e as rotinas de interface gráfica foram encapsuladas em um único pacote denominado **RBHTrans**. Dessa forma, todas as funcionalidades propostas são disponibilizadas a partir do comando:

```
R> require(RBHTrans)
```

Na sequência, o usuário pode acessar as interfaces gráficas do ambiente a partir dos comandos: `monta_banco()` e `escolhe_funcao()`.

4. Discussão e Conclusão

Este trabalho apresenta o método de varredura unidimensional desenvolvido especificamente para detecção de conglomerados de acidentes de trânsito em arruamentos. O método foi incorporado em um ambiente com interface gráfica que permite a análise dos eventos e dos conglomerados detectados utilizando recursos do *Google maps*. Dessa forma, o usuário pode visualizar remotamente o local do acidente de trânsito com grande riqueza de detalhes, além da disponibilidade de análises puramente espaciais. Trabalhos futuros tem como objetivo agregar informações de tráfego de veículos e pedestres na estimativa de intensidade de eventos, sob a hipótese nula.

Bibliografia

- Alm, S. E. (1997). On the distributions of scan statistics of a two dimensional Poisson process, *Advances in Applied Probability*, vol. 29, pages 1–18.
- Costa, M. A. and Kulldorff, M. (2009). In *Scan statistics: methods and applications*. *Birkhäuser: Statistics for Industry and Technology*, pages 129–52 [chapter 6].
- Costa, M. A. and Assunção, R. A. and Kulldorff, M. (2012). Constrained spanning tree algorithms for irregularly-shaped spatial clustering. *Computational Statistics and Data Analysis*. vol. 56, pages 1771–1783.
- Duczmal, L. and Assunção, R. A. (2004). Simulated annealing strategy for the detection of arbitrarily shaped spatial clusters, *Computational Statistics and Data Analysis*, vol. 45, pages 269–286.
- Gesmann, Markus and de Castillo, Diego (2011). Using the Google Visualisation API with R. *The R Journal*. vol. 3, n. 2, pages 40–44.
- Huang, L. and Stinchcomb, D. G. and Pickle, L. W. and Dill, J. (2009). Identifying clusters of active transportation using spatial scan statistics. *American Journal of Preventive Medicine*. vol. 37, n. 2, pages 157–166.
- Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods*, vol. 26, pages 1481–1496.
- Kulldorff, M. and Athas, W. and Feuer, E. and Miller, B. and Key, C. (1998). Evaluating cluster alarms: A space-time scan statistic and brain cancer in Los Alamos. *American Journal of Public Health*, vol. 88, pages 1377–1380.
- Kulldorff, M. (2001). Prospective time-periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society*, vol. A164, pages 61–72.
- Kulldorff, M. and Heffernan, R. and Hartman, J. and Assunção, R. M. and Mostashari, F. (2005). A space-time permutation scan statistic for the early detection of disease outbreaks. *PLoS Medicine*, vol. 2, pages 216–224.
- Kulldorff, M. and Huang, L. and Pickle, L. and Duczmal, L. (2006). An elliptic spatial scan statistic. *Statistics in Medicine*, vol. 25, pages 3929–3943.
- Loecher, Markus (2010). Plotting on Google Static Maps in R. Technical Report,