Estimativas de parâmetros de um modelo SAR a partir de métodos de otimização

José H C Monteiro da Silva

RA 117439

MI602 - 2s2020

1. Introdução

Este trabalho tem como objetivo estimar parâmetros de um modelo espacialmente autorregressivo (SAR) a partir do método de otimização de Newton-Rapshon.

A estratégia metodológica será aplicada aos dados de monitoramento de tiroteios nos bairros do município do Rio de Janeiro coletadas pela plataforma **Fogo Cruzado**[1]. Esse monitoramento é feito a partir do registro de ocorrências de tiroteios pelos próprios moradores utilizando um aplicativo de celular desenvolvido pela plataforma.

Este trabalho apresenta mais 3 seções além desta introdução. Na segunda seção serão apresentados os métodos de análise dos dados, ou seja, a descrição do modelo SAR aplicado aos dados de monitoramento de tiroteios e a metodologia de estimação dos parâmetros do modelo. Na terceira seção, apresentam-se os resultados das estimativas de parâmetros e os mapas de ocorrências suavizados a partir das estimativas geradas. Na última seção são apresentadas as referências bibliográficas do trabalho.

[1] Para maiores detalhes, consultar https://fogocruzado.org.br/

2. Métodos

2.1 Modelo Espacialmente Autorregressivo (SAR)

Para o presente trabalho, considera-se um modelo SAR sem covariáveis, dado por (CITAR):

$$Y_i - \mu = \rho \sum_{j \in v(i)} (Y_j - \mu) + \epsilon_i, \qquad (1)$$

onde,

- Y_i é a contagem de tiroteios no bairro i;
- μ é a média de tiroteios do universo de dados dos bairros;
- $v(i)$ é o conjunto de índices adjacentes à localização i, mas sem o próprio i, ou à grosso-modo, é o conjunto de bairros vizinhos ao bairro i;
- ρ é o parâmetro de autocorrelação espacial tal que $0 \leq \rho < 1$;
- ϵ_i é o erro atrelado à estimativa do bairro i com distribuição normal com média 0 e variância σ^2, ou seja, $\epsilon \sim N(0, \sigma^2 I)$.

Passando a equação (1) para o formato matricial, tem-se que:

$$Y - \mu\mathbf{1} = \rho A(Y - \mu\mathbf{1}) + \epsilon\mathbf{1}, \qquad (2)$$

em que:

- Y é o vetor n x 1 em que n é o número de bairros e cada elemento do vetor é a contagem de tiroteios;
- A é a matriz de adjacências construída de modo que $A_{ij} = 1$ caso o bairro i seja adjacente ao bairro j e 0, caso contrário, e $A_{ii} = 0$.

Rearranjando a equação (2), chega-se à equação (4).

$$(3) \qquad Y - \mu\mathbf{1} - \rho A(Y - \mu\mathbf{1}) = \epsilon\mathbf{1}$$

$$(4) \qquad (Y - \mu\mathbf{1})(I - \rho A) = \epsilon\mathbf{1}$$

Por fim, isolando Y, tem-se:

$$(5) \qquad Y = \mu\mathbf{1} + (I - \rho A)^{-1}\epsilon\mathbf{1},$$

assim, Y segue uma distribuição normal multivariada com média dada pelo vetor $\mu\mathbf{1}$ e desvio padrão $\sigma(I - \rho A)^{-1}$:

$$(6) \qquad Y \sim N(\mu\mathbf{1}, \sigma^2(I - \rho A)^{-2}).$$

A partir da distribuição de Y, pode-se encontrar um estimador de máxima verossimilhança para os parâmetros μ, $sigma^2$ e ρ, tal que:

$$(7) \qquad L(\mu, \rho, \sigma^2) = \frac{|\sigma^2(I - \rho A)^{-2}|^{-n/2}}{(2\pi)^{n/2}} \exp\Big[-\frac{1}{2}(Y - \mu\mathbf{1})^T(\sigma^2(I - \rho A)^{-2})^{-1}(Y - \mu\mathbf{1})\Big],$$

dessa forma, pode-se estimar μ, σ^2 e ρ a partir da maximização do logarítimo da função de verossimilhança $L(\mu, \rho, \sigma^2) = \log(L(\mu, \rho, \sigma^2)) = l(\mu, \rho, \sigma^2)$, o que transforma o problema de estimação dos parâmetros em um problema de otimização. Para sua solução, pode-se lançar mão, por exemplo, do algoritmo de Newton-Rapshon, apresentado a seguir.

2.2 Método de Newton-Raphson

Partindo da função $l(\mu, \rho, \sigma^2)$ definida anteriormente como o logarítimo da função de verossimilhança para o modelo SAR adotado para o problema, utiliza-se do Método de Newton-Raphson para a estimativa dos parâmetros $\hat{\mu}, \hat{\sigma^2}, \hat{\rho}$. O método consiste do uso da expansão de Taylor com o objetivo de aproximar raízes de funções, podendo-se adaptadas para estimar seus máximos e mínimos (Ludwig 2020).

No caso de funções com mais de um parâmetro no domínio dos reais, são utilizados dos gradientes e da matriz Hessiana para delimitação dos passos de otimização. Para o exemplo do presente trabalho, definem-se:

- $p_k = (\mu_k, \sigma_k^2, \rho_k)^T$, vetor de parâmetros no passo k do método de Newton-Raphson;
- $\nabla l(p)$, vetor gradiente do logarítimo da função de verossimilhança, dado por:

$$\nabla l(\mu, \sigma^2, \rho) = \begin{bmatrix} \frac{\partial l}{\partial \mu} \\ \frac{\partial l}{\partial \sigma^2} \\ \frac{\partial l}{\partial \rho} \end{bmatrix}; \tag{8}$$

- $\mathbf{H}_l(p)$, matriz Hessiana do logarítimo da função de verossimilhança, dada por:

$$\mathbf{H}_{l(\mu,\sigma^2,\rho)} = \begin{bmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \mu \partial \sigma^2} & \frac{\partial^2 l}{\partial \mu \partial \rho} \\ \frac{\partial^2 l}{\partial \sigma^2 \partial \mu} & \frac{\partial^2 l}{\partial (\sigma^2)^2} & \frac{\partial^2 l}{\partial \sigma^2 \partial \rho} \\ \frac{\partial^2 l}{\partial \rho \partial \mu} & \frac{\partial^2 l}{\partial \rho \partial \sigma^2} & \frac{\partial^2 l}{\partial \rho^2} \end{bmatrix} \tag{9}$$

Assim, partindo de um chute inicial dado por $k = 0$, definido por $p_0 = (\mu_0, \sigma_0^2, \rho_0)^T$, define-se a aproximação para a estimativa desejada de p_k a partir do seguinte passo:

$$p_k = p_{k-1} - \left[\mathbf{H}_l(p_{k-1})\right]^{-1} \nabla l(p_{k-1}) \tag{10}$$

No presente exercício, utilizou-se da função *nlm (Non-Linear Minimization)* do *software* **R** para gerar as estimativas dos parâmetros desejados. Serão apresentados na seção a seguir os diferentes cenários escolhidos de partida p_0 e seus resultados de estimativas finais de modo a discutir a dificuldade de convergência e possíveis distorções que determinadas entradas podem gerar no método Newton-Raphson.

3. Resultados

3.1 Análise Descritiva

O mapa da Figura 1 apresenta a disposição espacial da contagem de tireoteios nos bairros do Rio de Janeiro. No período de análise, observou-se uma média de 9.52 tiroteios por bairro da cidade com desvio padrão de 15.03. O maior número de tiroteios no período foi registrado na Cidade de Deus com 101 registros, seguido do Complexo do Alemão com 86 e da Tijuca com 66. Por outro lado, 34 dos 159 bairros analisados não registraram tiroteios, dentre os quais a Gávea, a Lapa e São Conrado. Notadamente, os bairros da Zona Sul e região central da cidade são os que apresentam menores números de registros.

3.2 Estratégia de otimização

A estimação dos parâmetros via método de Newton-Raphson envolve a definição dos chutes iniciais para μ_0, σ_0 e rho_0. Para gerar uma grande quantidade de amostras e testar a convergência do método para diferentes *inputs*, criaram-se 5 conjuntos de chutes iniciais, apresentados na Tabela 1. Manteve-se μ_0 constante em 4 listas de chutes à exceção da lista 4 em que se criaram conjuntos com σ_0 e ρ_0 fixos e μ_0 variando em passos de 5. O mesmo foi feito para a lista 5 no caso do desvio-padrão. Nas listas 1, 2 e 3 buscaram-se variações de ρ_0 desde 0.01 a 0.95 em passos de 0.01.

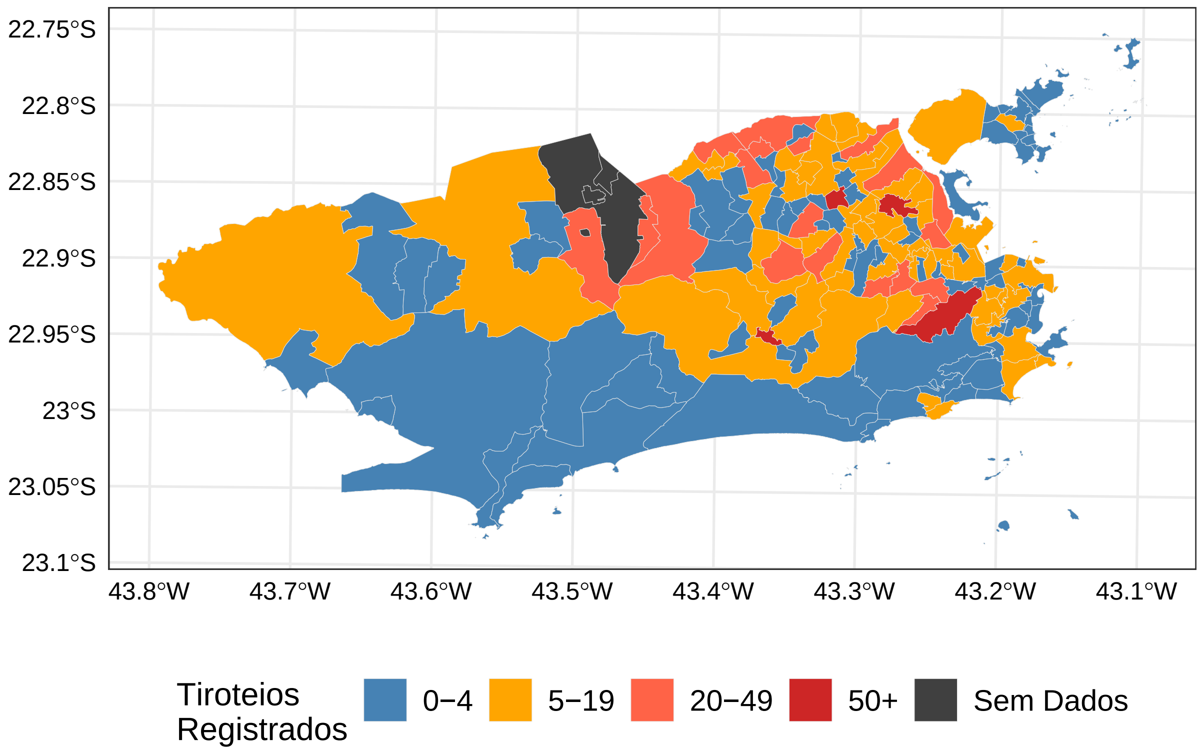

Figura 1: Número de tiroteios registrados nos bairros do Rio de Janeiro no período analisado. Fonte: Plataforma Fogo Cruzado, 2020.

Tabela 1: Critérios para criação de listas para chutes iniciais para estimação de parâmetros via Método de Newton-Raphson.

Lista	μ_0	σ_0	ρ_0
1	9.52	15.03	0.01-0.95
2	25	25	0.01-0.95
3	50	2	0.01-0.95
4	5-200	15.03	0.01
5	10	5-200	0.01

Na Tabela 2 apresentm-se um conjunto de amostras de resultados obtidos a partir de 10 diferentes critérios de chute inicial.

Tabela 2: Resultados de estimativas dos parâmetros do modelo SAR via Método de Newton-Raphson.

(μ_0,σ_0,ρ_0)	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}$	Iterações
(9.52, 15.03, 0.01)	9.36	14.96	0.012	19
(9.52, 15.03, 0.80)	9.75	17.54	0.259	35
(25, 25, 0.04)	9.36	14.96	0.012	34
(25, 25, 0.53)	11.02	24.57	0.535	56
(50, 2, 0.18)	5.43	15.82	0.162	76
(50, 2, 0.04)	9.36	14.96	0.012	65
(50, 100, 0.01)	9.36	14.96	0.012	41
(25, 100, 0.01)	11.05	32.53	0.786	31
(10, 75, 0.01)	10.65	19.45	-0.313	29
(10, 5, 0.01)	9.36	14.96	0.012	25

Referências Bibliográficas

Ludwig, G. (2020), “Notas de Aula - MI602 Métodos Computacionais em Estatística,” Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas.

Agradecimentos

Agradeço à colega Marília Gabriela Rocha pelo suporte durante a realização dessa atividade e, principalmente, pela paciência com as dúvidas elementares.

Material para Replicação

Os códigos utilizados no presente trabalho encontram-se disponíveis no repositório do **github**: https://github.com/josehcms/MI602_2s2020.