

– quantil –

Predicción de Crimen en Bogotá

Juan Felipe Cerón Uribe

30 de marzo del 2017

Fuente principal:

G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, y G. E. Tita. "Self-exciting point process modeling of crime". En: *Journal of the American Statistical Association* (2011).

Contenido

- 1 Descripción del problema
- 2 Modelamiento
 - Procesos puntuales
 - Modelo ETAS
- 3 Estimación de distribuciones
- 4 Metodología
- 5 Validación de modelos

Descripción del problema

- El objetivo es generar un modelo predictivo del crimen en Bogotá.

Descripción del problema

- El objetivo es generar un modelo predictivo del crimen en Bogotá.
- Queremos determinar la intensidad relativa del crimen en puntos espacio-temporales.

Descripción del problema

- El objetivo es generar un modelo predictivo del crimen en Bogotá.
- Queremos determinar la intensidad relativa del crimen en puntos espacio-temporales.
- Contamos con datos de 300k crímenes entre 2004-2016 (entre otras).

Modelamiento

Procesos puntuales

Un **proceso puntual** es una colección aleatoria de puntos cayendo sobre un espacio [6].

- En nuestro caso son puntos espacio-temporales.

Procesos puntuales

Un **proceso puntual** es una colección aleatoria de puntos cayendo sobre un espacio [6].

- En nuestro caso son puntos espacio-temporales.
- Sea $N\{(t_1, t_2) \times (x_1, x_2) \times (y_1, y_2)\}$ el número de eventos ocurridos en esa caja (es una variable aleatoria).

Queremos aproximar la **intensidad condicional** del proceso:

$$\lambda(t, x, y) = \lim_{\Delta t, \Delta x, \Delta y \rightarrow 0} \frac{E[N\{(t + \Delta t) \times (x + \Delta x) \times (y + \Delta y)\}]}{(t + \Delta t)(x + \Delta x)(y + \Delta y)}$$

$$= \lim_{\text{región} \rightarrow 0} \frac{\# \text{ eventos esperados en la región}}{\text{tamaño de la región}}$$

Queremos aproximar la **intensidad condicional** del proceso:

$$\lambda(t, x, y) = \lim_{\Delta t, \Delta x, \Delta y \rightarrow 0} \frac{E[N\{(t + \Delta t) \times (x + \Delta x) \times (y + \Delta y)\}]}{(t + \Delta t)(x + \Delta x)(y + \Delta y)}$$

$$= \lim_{\text{región} \rightarrow 0} \frac{\# \text{ eventos esperados en la región}}{\text{tamaño de la región}}$$

- Análoga a una densidad de probabilidad.

Modelo ETAS

Modelo **Epidemic Type Aftershock Sequence**:

$$\lambda(t, x, y) = \mu(t, x, y) + \sum_{\{k: t_k < t\}} g(t - t_k, x - x_k, y - y_k)$$

Modelo ETAS

Modelo **Epidemic Type Aftershock Sequence**:

$$\lambda(t, x, y) = \mu(t, x, y) + \sum_{\{k:t_k < t\}} g(t - t_k, x - x_k, y - y_k)$$

- La intensidad puntual tiene un componente de trans fondo y uno dependiente de la historia.

Modelo ETAS

Modelo **Epidemic Type Aftershock Sequence**:

$$\lambda(t, x, y) = \mu(t, x, y) + \sum_{\{k:t_k < t\}} g(t - t_k, x - x_k, y - y_k)$$

- La intensidad puntual tiene un componente de trans fondo y uno dependiente de la historia.
- Suponemos que el proceso es **auto excitante**: La ocurrencia de eventos incrementa la probabilidad de otros en el futuro.

Justificación

Varias fuentes sugieren que el fenómeno del crimen se estimula a si mismo:

- Investigaciones de criminalística muestran que puede esparcirse localmente como un proceso contagioso [4].
- Ladrones atacan repetidamente clusters de objetivos cercanos pues conocen vulnerabilidades locales [7].
- Un evento de violencia entre pandillas puede incitar oleadas de violencia como venganza [3].
- La evidencia indica un riesgo elevado de hurto en la vecindad de una residencia atacada recientemente [2].

La forma paramétrica de estas funciones en sismología lleva años de desarrollo. ¿Cómo agilizamos este proceso para la estimación del crimen?

Estimación de distribuciones

Un **kernel** es una función real no negativa que satisface:

- $\int_{\mathbb{R}} K(u) du = 1$
- $K(-u) = K(u)$

Estimación de distribuciones

Un **kernel** es una función real no negativa que satisface:

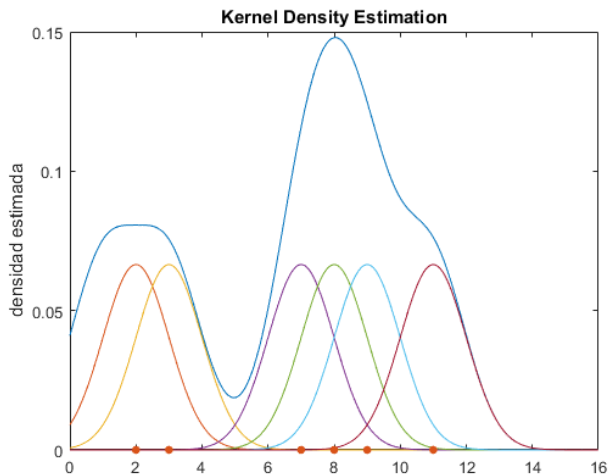
- $\int_{\mathbb{R}} K(u) du = 1$
- $K(-u) = K(u)$

Sean $\{x_i\}_{i \leq n}$ iid, $h \in \mathbb{R}$. El **estimador de densidad de kernels** de los datos es:

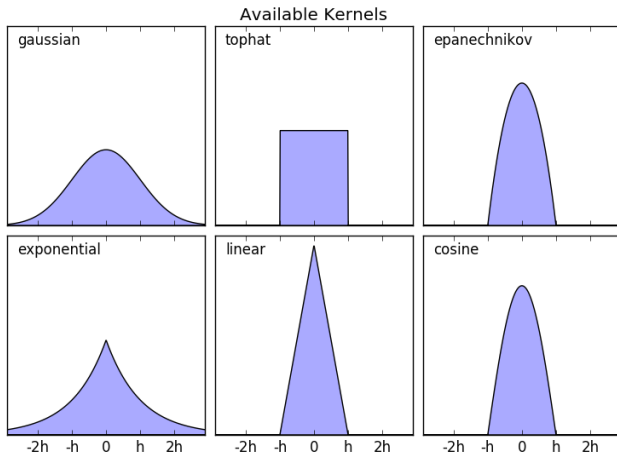
$$\hat{p}_h(x) = \frac{1}{n} \sum_{i=0}^n \frac{1}{h} K\left(\frac{x - x_i}{h}\right)$$

El parámetro h se conoce como **ancho de banda**.

Cada kernel esparce $\frac{1}{n}$ de masa de probabilidad:



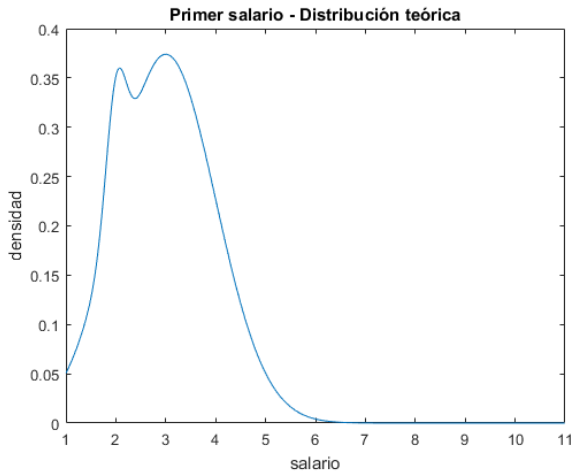
Algunas opciones de kernels



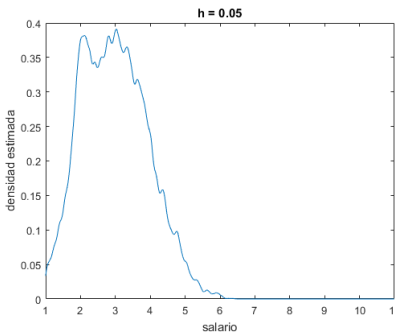
- La mayoría de las densidades unimodales tienen prácticamente el mismo desempeño como kernels [5].
- El parámetro h tiene un alto impacto en la predicción.

Impacto del ancho de banda

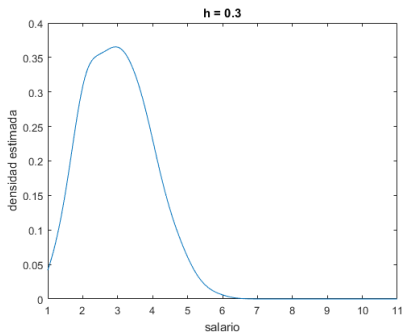
Ejemplo



Impacto del ancho de banda

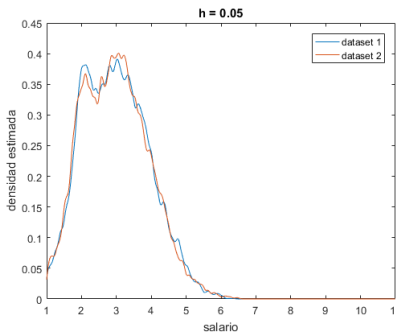


(a) h demasiado bajo

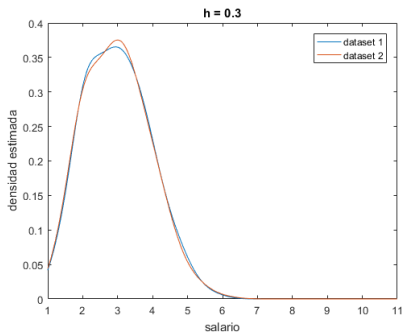


(b) h demasiado alto

Sesgo vs varianza



(a) varianza alta



(b) sesgo alto

Sesgo y varianza

Suponga la siguiente relación:

$$Y = f(X) + \epsilon$$

donde $E[\epsilon] = 0$ es un error aleatorio.

Sesgo y varianza

Suponga la siguiente relación:

$$Y = f(X) + \epsilon$$

donde $E[\epsilon] = 0$ es un error aleatorio. Estimamos Y con $\hat{f}(X)$. El sesgo y la varianza de la estimación son:

$$\text{sesgo}(\hat{f}(X)) = E \left[\hat{f}(X) - f(X) \right]$$

$$\text{var}(\hat{f}(X)) = E \left[(\hat{f}(X) - E[\hat{f}(X)])^2 \right]$$

Sesgo y varianza

El error de estimación está dado por:

$$\begin{aligned} E \left[(Y - \hat{f}(X))^2 \right] &= E \left[(f(X) - \hat{f}(X))^2 \right] + \text{var}(\epsilon) \\ &= \text{sesgo}^2(\hat{f}(X)) + \text{var}(\hat{f}(X)) + \text{var}(\epsilon) \end{aligned}$$

Selección heurística del ancho de banda

Elegir el ancho de banda es difícil, pero existen buenas alternativas heurísticas.

Selección heurística del ancho de banda

Elegir el ancho de banda es difícil, pero existen buenas alternativas heurísticas. Sean D_i el m –ésimo vecino más cercano al evento i , σ_t la desviación estándar de los tiempos:

$$\hat{v}(t) = \frac{1}{n} \sum_{i=0}^n \frac{1}{\sigma_t D_i} K \left(\frac{t - t_i}{\sigma_t D_i} \right)$$

Selección heurística del ancho de banda

Elegir el ancho de banda es difícil, pero existen buenas alternativas heurísticas. Sean D_i el m –ésimo vecino más cercano al evento i , σ_t la desviación estándar de los tiempos:

$$\hat{\nu}(t) = \frac{1}{n} \sum_{i=0}^n \frac{1}{\sigma_t D_i} K \left(\frac{t - t_i}{\sigma_t D_i} \right)$$

$$\hat{\mu}(t, x, y) = \frac{1}{n} \sum_{i=0}^n \frac{1}{\sigma_t \sigma_x \sigma_y D_i^3} K \left(\frac{t - t_i}{\sigma_t D_i} \right) K \left(\frac{x - x_i}{\sigma_x D_i} \right) K \left(\frac{y - y_i}{\sigma_y D_i} \right)$$

Metodología

Declustering estocástico

Es un algoritmo que pretende separar los eventos de transfondo de las réplicas para estimar cada componente de la intensidad mediante kernels.

Suponga que tenemos una base de datos de N eventos. Sea P la matriz $N \times N$ tal que p_{ij} es la probabilidad de que el evento j causara al evento i . Si el modelo es correcto tenemos las siguientes relaciones:

$$p_{ii} = \frac{\mu(t_i, x_i, y_i)}{\lambda(t_i, x_i, y_i)}$$

$$p_{ji} = \frac{g(t_i - t_j, x_i - x_j, y_i - y_j)}{\lambda(t_i, x_i, y_i)}$$

Algoritmo

Iniciamos con una matriz P_0 aleatoria e iteramos hasta la convergencia:

Algoritmo

Iniciamos con una matriz P_0 aleatoria e iteramos hasta la convergencia:

- 1 Usando P_{n-1} simulamos cuáles eventos $\{(t_i, x_i, y_i)\}_{i=0}^{N_t}$ son de transfondo y obtenemos las distancias $\{(dt_i, dx_i, dy_i)\}_{i=0}^{N_r}$ de las réplicas a sus padres. $N_t + N_r = N$.

Algoritmo

Iniciamos con una matriz P_0 aleatoria e iteramos hasta la convergencia:

- 1 Usando P_{n-1} simulamos cuáles eventos $\{(t_i, x_i, y_i)\}_{i=0}^{N_t}$ son de transfondo y obtenemos las distancias $\{(dt_i, dx_i, dy_i)\}_{i=0}^{N_r}$ de las réplicas a sus padres. $N_t + N_r = N$.
- 2 Estimamos μ_n de los eventos de transfondo y g_n de las distancias.

Algoritmo

Iniciamos con una matriz P_0 aleatoria e iteramos hasta la convergencia:

- 1 Usando P_{n-1} simulamos cuáles eventos $\{(t_i, x_i, y_i)\}_{i=0}^{N_t}$ son de transfondo y obtenemos las distancias $\{(dt_i, dx_i, dy_i)\}_{i=0}^{N_r}$ de las réplicas a sus padres. $N_t + N_r = N$.
- 2 Estimamos μ_n de los eventos de transfondo y g_n de las distancias.
- 3 Utilizamos las ecuaciones del anterior slide para actualizar P_n .

Resultados del autor

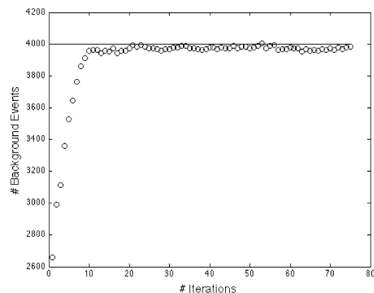
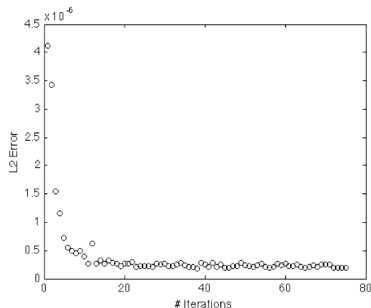


Figura: resultados en *Self-exciting point process modelling of crime*

Análisis del algoritmo

El algoritmo busca agrupamientos espacio-temporales de puntos (para estimar μ) y de distancias padre-réplica (para estimar g):

Análisis del algoritmo

El algoritmo busca agrupamientos espacio-temporales de puntos (para estimar μ) y de distancias padre-réplica (para estimar g):

- 1 Un evento se clasifica como de transfondo.

Análisis del algoritmo

El algoritmo busca agrupamientos espacio-temporales de puntos (para estimar μ) y de distancias padre-réplica (para estimar g):

- 1 Un evento se clasifica como de transfondo.
- 2 μ se eleva cerca de ese evento.

Análisis del algoritmo

El algoritmo busca agrupamientos espacio-temporales de puntos (para estimar μ) y de distancias padre-réplica (para estimar g):

- 1 Un evento se clasifica como de transfondo.
- 2 μ se eleva cerca de ese evento.
- 3 En la siguiente iteración incrementa la probabilidad de que un evento vecino sea de transfondo.

Análisis del algoritmo

El algoritmo busca agrupamientos espacio-temporales de puntos (para estimar μ) y de distancias padre-réplica (para estimar g):

- 1 Un evento se clasifica como de transfondo.
- 2 μ se eleva cerca de ese evento.
- 3 En la siguiente iteración incrementa la probabilidad de que un evento vecino sea de transfondo.
- 4 Si son muchos, es probable que varios se clasifiquen como de transfondo, identificando un cluster.

Análisis del algoritmo

La matriz P deja de oscilar cuando una nueva simulación no altera mucho el estado de las funciones μ y g .

- Esto sucede cuando las intensidades reflejan la concentración de puntos y de distancias respectivamente.

Validación de modelos

Simulación de un proceso ETAS

Podemos simular un proceso como el visto dadas funciones μ y g . Después de ejecutar el algoritmo podemos analizar:

- El número de eventos de transfondo teórico contra el aproximado.
- El error integrado en una región:

$$e = \int_B (\lambda(u) - \hat{\lambda}(u))^2 du$$

Análisis residual

Definimos el **residuo** de la aproximación en una región observada B como:

$$R(B) = N_B - \int_B \hat{\lambda}(u) du$$

donde N_B es el número de eventos observados en la región. Esta medida permite comparar la bondad de ajuste de varios modelos [1].

Análisis residual

Definimos el **residuo** de la aproximación en una región observada B como:

$$R(B) = N_B - \int_B \hat{\lambda}(u) du$$

donde N_B es el número de eventos observados en la región. Esta medida permite comparar la bondad de ajuste de varios modelos [1].

- Un caso interesante es cuando B no hizo parte de los datos de entrenamiento.

Teorema del re-escalamiento temporal








Sean $u_1, u_2, \dots, u_n < T$ realizaciones de un proceso puntual con intensidad condicional λ . Defina la transformación:

$$A(u_k) = \int_0^{u_k} \lambda(u|H_u) du$$

Los $A(u_k)$'s son un proceso de Poisson con parámetro 1. El tiempo entre eventos de un proceso de Poisson tiene una distribución exponencial, por lo cuál se presta para pruebas de hipótesis.

A futuro

- Incorporar nuevas variables al modelo (hora del día, día de la semana, tipo de crimen).
- Evitar que la función de generación de réplicas g propague una densidad hacia el pasado.
- Aprender más sobre anchos de banda.
- Experimentar con modelos diferentes (incluyendo paramétricos).
- Desarrollo de una herramienta de comparación de modelos.
- Implementar exitosamente para el progreso de la sociedad.

-  J. Moller M. Hazelton A. Baddeley R. Turner. “Residual analysis for spatial point processes”. En: *Journal of the Royal Statistical Society* (2005).
-  K. Pease G. Farrell. “Repeat victimization”. En: *New York: Criminal Justice Press* (2001).
-  G. Ridgeway G. Tita. “The impact of gang formation on local patterns of crime”. En: *Journal of Research on Crime and Delinquency* (2007).
-  S.D. Johnson. “Repeat burglary victimisation: A tale of two theories”. En: *Journal of Experimental Criminology* (2008).
-  M.C. Jones M.P. Wand. *Kernel Smoothing*. Chapman y Hall, 1995.
-  Frederic Paik Schoenberg. “Introduction to point processes”. En: ().
-  P. Nieuwbeerta W. Bernasco. “How do residential burglars select target areas? A new approach to the analysis of criminal location choice”. En: *The British Journal of Criminology* (2005).

GRACIAS