

Spatio-temporal methods for the analysis of crime and traffic safety data

Álvaro Briz-Redón

Departament d'Estadística i Investigació Operativa

Universitat de València

ORCID: 0000-0001-7976-3534

alvaro.briz@uv.es

Directores:

Francisco Montes Suay (Universitat de València)

Francisco Martínez Ruiz (Ajuntament de València)

Keywords: spatial statistics, spatio-temporal statistics, statistical modeling, geographic information science, applied statistics.

MSC Subject classifications: 62H11, 62M30, 62P99.

En 1855, el médico inglés John Snow presentó un mapa que mostraba la ubicación de los casos de cólera detectados durante la epidemia londinense de 1854. Snow investigó el patrón formado por estas ubicaciones y sugirió que una fuente de agua pública situada en Broad Street había sido la causante del origen del brote epidémico. La creación de este mapa y el posterior estudio de la evolución de la epidemia en base al mismo suele considerarse como el primer caso de análisis estadístico espacial. Desde entonces, han surgido numerosos ejemplos de datos espaciales, especialmente durante las últimas décadas, con aplicaciones diversas en múltiples áreas científicas tales como la agricultura, la astronomía, la biología, la epidemiología, la hidrología, la meteorología y la teledetección, entre otras. En concreto, existen tres tipos principales de datos espaciales: los *patrones puntuales*, los *datos espacialmente agregados* y los *datos geoestadísticos*. Cada uno de estos tipos ha dado lugar al desarrollo de sendas líneas de investigación.

En nuestro caso, la colaboración existente entre la Policía Local de València y el Departamento de Estadística e Investigación Operativa de la Universitat de València facilitó que la tesis doctoral que aquí se resume (Briz Redón, 2020) se orientara hacia el estudio de problemas de naturaleza espacial y/o espacio-temporal de especial interés en los campos de la criminología y la seguridad vial, y hacia el uso y desarrollo de métodos estadísticos adecuados para su análisis. Desde un punto de vista metodológico, se trataron técnicas propias del análisis de patrones puntuales y de modelización de datos agregados con dependencia espacial. En particular, durante la primera fase de la tesis (lo que corresponde fundamentalmente a sus capítulos 2 y 3) se emplearon estructuras espaciales de tipo grafo (redes lineales) correspondientes a la red viaria de València como soporte de los

datos. Este hecho, necesario para el adecuado análisis de eventos que ocurren sobre la red viaria, conlleva ciertas dificultades técnicas y metodológicas que fueron abordadas. En los siguientes capítulos (capítulos 4 a 7) se estudiaron diversos problemas de índole espacial o espacio-temporal de gran interés en criminología y seguridad vial haciendo uso, en general, de datos espacialmente agregados sobre unidades geográficas dotadas de área. En los siguientes párrafos se desglosan brevemente los contenidos de la tesis siguiendo el orden de sus capítulos.

En el Capítulo 2 se presenta una modelización espacial de los conteos de accidentes de tráfico a nivel de calle sobre la red viaria de un distrito de València, con el fin de determinar factores de riesgo. En particular, debido a la alta presencia de accidentes alrededor de los puntos de intersección entre calles, la red viaria original se segmentó de modo que permitiese distinguir los tramos de calle de tipo intersección de aquellos que quedan entre estos primeros. Además, la modelización de los conteos se complementó con la detección de *coldspots* y *hotspots* (tramos de calle de baja y alta concentración de accidentes, respectivamente) mediante técnicas no paramétricas para la estimación de la intensidad de un proceso puntual sobre una red lineal, combinados con el uso de indicadores locales de asociación espacial. Los resultados asociados a este análisis pueden consultarse en Briz-Redón, Martínez-Ruiz y Montes (2019d).

El Capítulo 3 se centró en la detección de *hotspots* sobre la red viaria de València donde un tipo de accidente de tráfico está sobrerrepresentado. Así pues, se puso el foco en el riesgo relativo que supone una vía para un tipo de accidente de especial relevancia (como, por ejemplo, el atropello) y no en la concentración total del mismo. Se adaptaron algunas metodologías no paramétricas previamente propuestas para la construcción de superficies de riesgo para estimar la probabilidad relativa de ocurrencia de un tipo concreto de evento a lo largo de una red lineal. La Figura 1 muestra un ejemplo de estimación de esta probabilidad relativa sobre la red de València. A continuación, se propuso una metodología para detectar *hotspots* de riesgo diferencial en la red de acuerdo a los valores de probabilidad relativa previamente inferidos, teniendo en cuenta el tamaño muestral asociado a los valores de probabilidad estimados. Esta metodología se describe detalladamente en Briz-Redón, Martínez-Ruiz y Montes (2019b), aplicándose sobre un patrón marcado de accidentes de tráfico que incluía información sobre el tipo de colisión y sobre los tipos de vehículos implicados en cada accidente. El paquete de R disponible en CRAN llamado DRHotNet (*Differential Risk Hotspots in a Linear Network*) contiene la implementación completa de este método. Este paquete se describe en profundidad en Briz-Redón, Martínez-Ruiz y Montes (2021).

El Capítulo 4 de la tesis se dedica al fenómeno de la repetición en criminología, siendo el principal análisis espacio-temporal incluido en la tesis. La mayoría de delitos han sido investigados desde la perspectiva de la repetición o casi-repetición, conceptos que hacen referencia a la mayor probabilidad de observar un delito en la proximidad (en espacio y tiempo) de un evento delictivo previo. Por tanto, este análisis se basa en asumir que el crimen presenta un comportamiento similar al de un proceso contagioso o epidémico. De hecho, el test de Knox, la herramienta estadística más habitual para evaluar la magnitud y extensión espacio-temporal del fenómeno de la casi-repetición, fue concebido originalmente para estudios epidemiológicos. El principal inconveniente del test de Knox es que no tiene en cuenta la heterogeneidad espacio-temporal del riesgo de crimen, lo que en el contexto de la repetición de delitos complica la distinción entre las casi-repeticiones que están relacionadas (explicadas por la teoría criminológica conocida como *boost*) y las que no lo están (explicadas por la llamada teoría *flag*). En este capítulo se ajusta la versión clásica del test de Knox para analizar la magnitud y extensión del fenómeno de casi-repetición en un conjunto de datos de robos a vivienda registrados en València. Esta aportación metodológica se presentó en Briz-Redón, Martínez-Ruiz y Montes (2022).



Fig. 1: Estimación de la probabilidad relativa de accidente para una motocicleta sobre la red viaria de los distritos centrales de València

El Capítulo 5 se centró en la modelización del número de accidentes de circulación sobre diferentes particiones de la ciudad de València basadas en su estructura administrativa o de tráfico. En este sentido, es sabido que la elección de una determinada unidad de análisis territorial puede afectar a los resultados del método estadístico que se emplee. Esta cuestión se conoce como el problema de la unidad espacial modificable (MAUP), término que fue introducido por el geógrafo S. Openshaw a finales de los años 70 del siglo pasado. Este capítulo incluye un estudio de caso sobre la ciudad de València en el que se analizan las consecuencias de variar la escala o la zonificación que se escoge para llevar a cabo la modelización de los conteos de accidentes. En concreto, se investigó el efecto que tiene modificar cualquiera de estos dos factores sobre la distribución de la variable respuesta y las covariables, y el impacto en la estimación de los efectos de las covariables, considerando el modelo espacial de tipo condicional autorregresivo y la regresión geográficamente ponderada. El estudio completo puede encontrarse en Briz-Redón, Martínez-Ruiz y Montes (2019c).

En la literatura criminológica se distinguen tres tipos principales de lugares en relación con su potencial conexión con el crimen: los que atraen el crimen, los que favorecen indirectamente su generación y los que lo previenen. En general, existen múltiples alternativas de modelización estadística para evaluar si cierta localización (o conjunto de localizaciones) de una ciudad influye en la incidencia de un evento de interés, algunas de las cuales surgieron en un contexto epidemiológico con el fin de estudiar la posible relación entre la presencia de una fuente de contaminación atmosférica y la incidencia de ciertos tipos de cáncer. En el Capítulo 6 de la tesis se comparan diversos métodos estadísticos que permiten abordar este problema, con el fin secundario de analizar si existe relación entre la intensidad de los accidentes de tráfico en València y la localización de los colegios e institutos de la ciudad. El análisis comparativo completo se encuentra en Briz-Redón, Martínez-Ruiz y Montes (2019a)

El último estudio que se llevó a cabo durante el desarrollo de esta tesis trató sobre los efectos de la existencia de datos faltantes en un contexto criminológico como consecuencia de la no geocodificación

de ciertos eventos. Este tema es de interés en cualquier disciplina que requiera del uso de datos espaciales, pero ha sido particularmente enfatizado en el ámbito de la criminología cuantitativa. En 2004, el criminólogo J. Ratcliffe sugirió que una tasa de geocodificación del 85 % podía considerarse aceptable para llevar a cabo un análisis espacial de un tipo de crimen. El Capítulo 7 contiene una reestimación de esta tasa teniendo en cuenta algunos factores espaciales (niveles de intensidad, clusterización y agregación) que se pasaron por alto en dicho estudio previo. Además, el procedimiento propuesto por Ratcliffe, basado en el clásico test de Mann-Whitney, se extendió a través de otros métodos estadísticos que resultan más adecuados para el tratamiento de datos espaciales. Así, se demostró que la tasa de geocodificación del 85 % inicialmente propuesta puede ser insuficiente bajo determinadas condiciones. Estos resultados fueron publicados en Briz-Redón, Martínez-Ruiz y Montes (2020).

Agradecimientos

A la Policía Nacional, a la Policía Local de València y a la Oficina de Estadística del Ayuntamiento de València, por proporcionar los datos que se emplearon para la realización de esta tesis. A mis directores de tesis, por toda la ayuda prestada.

Acerca del autor



Álvaro Briz Redón es Licenciado en Matemáticas (2014) y Doctor en Estadística y Optimización (2020) por la Universitat de València. Actualmente es Profesor Ayudante Doctor en el Departamento de Estadística e Investigación Operativa de la misma universidad. Su investigación se enmarca en el área de la estadística espacio-temporal y sus aplicaciones.

Referencias

- Briz Redón, Á (2020). «Spatio-temporal methods for the analysis of crime and traffic safety data». Tesis doct. Universitat de València. URL: <https://roderic.uv.es/handle/10550/75234>.
- Briz-Redón, Álvaro, Francisco Martínez-Ruiz y Francisco Montes (2019a). «Estimating the occurrence of traffic accidents near school locations: A case study from Valencia (Spain) including several approaches». En: *Accident Analysis & Prevention* 132, pág. 105237.
- (2019b). «Identification of differential risk hotspots for collision and vehicle type in a directed linear network». En: *Accident Analysis & Prevention* 132, pág. 105278.
- (2019c). «Investigation of the consequences of the modifiable areal unit problem in macroscopic traffic safety analysis: a case study accounting for scale and zoning». En: *Accident Analysis & Prevention* 132, pág. 105276.
- (2019d). «Spatial analysis of traffic accidents near and between road intersections in a directed linear network». En: *Accident Analysis & Prevention* 132, pág. 105252.
- (2020). «Reestimating a minimum acceptable geocoding hit rate for conducting a spatial analysis». En: *International Journal of Geographical Information Science* 34.7, págs. 1283-1305.
-

- Briz-Redón, Álvaro, Francisco Martínez-Ruiz y Francisco Montes (2021). «DRHotNet: An R package for detecting differential risk hotspots on a linear network». En: *R Journal* 13.2, págs. 380-397.
- (2022). «Adjusting the Knox test by accounting for spatio-temporal crime risk heterogeneity to analyse near-repeats». En: *European Journal of Criminology* 19.4, págs. 586-611.
-