

RECONSTRUYENDO LOS COMPORTAMIENTOS INDIVIDUALES A PARTIR DE DATOS AGREGADOS: LPHOM

José M. Pavía

Universitat de Valencia

<https://orcid.org/0000-0002-0129-726X>

<https://www.researchgate.net/profile/Jose-Pavia-2>

pavia@uv.es

Proyecto de Investigación: AICO/2021/257

Abstract

La inferencia ecológica tiene como objeto estimar los valores de las celdas internas de un conjunto de tablas de contingencia $R \times C$ relacionadas cuando solo se conocen los agregados marginales. Esto define uno de los problemas cuantitativos más antiguos y con más lustre de las ciencias sociales (King, 1997), presente en muchas áreas, incluidas la economía, la epidemiología y la historia cuantitativa, aunque la ciencia política y la sociología son las disciplinas en las que aparece con mayor frecuencia. Ejemplos típicos lo constituyen la estimación de las transferencias de voto entre elecciones o discernir como se vota en función de la clase social. Esta información suele resumirse mediante una matriz de proporciones (probabilidades) filas estandarizadas. Las encuestas constituyen un medio, no exento de problemas y limitaciones, para obtener estimaciones de tales matrices. Los métodos de inferencia ecológica (solos o en combinación con otras fuentes) constituyen otra alternativa. En los últimos años hemos vivido una explosión de métodos para resolver estos problemas desde la estadística Bayesiana, prácticamente todos ellos sustentados en un modelo jerárquico multinomial Dirichlet Bayesiano (Rosen et al., 2001). El empleo de esta metodología (Olivia et al., 2020), sin embargo, requiere de analistas altamente entrenados y suele conllevar unos costes computacionales elevados (Romero y Pavía, 2021; Pavía y Romero, 2022c). Recientemente ha aparecido una nueva metodología, sustentada en programación matemática, que simplifica notablemente la resolución de estos problemas (Pavía y Romero, 2022a), hasta convertirlos en algo casi mecánico. Una parte importante de los nuevos algoritmos están disponibles en el paquete lphom (Pavía y Romero, 2022b) del software estadístico R. El objetivo de esta comunicación es exponer las potencialidades que presenta esta nueva aproximación, mostrando su uso a través de diversos ejemplos y profundizando en las mejoras de precisión que adelantan las nuevas extensiones basadas en:

- (i) la introducción de información a priori, (ii) la clusterización (automática) de unidades o
- (iii) el uso de algoritmos de inteligencia artificial.

Palabras clave: Matrices origen-destino; lphom; programación lineal; sociología política; aprendizaje automático