Disease risk estimation in small areas accounting for local spatial and spatio-temporal discontinuities

Guzmán Santafé[1], *Aritz Adin*[2], *María Dolores Ugarte*[3]

[1]guzman.santafe@unavarra.com, Department of Statistics, Computer Science and Mathematics, Public University of Navarre

[2]aritz.adin@unavarra.es, Department of Statistics, Computer Science and Mathematics, Public University of Navarre

[3]lola@unavarra.es, Department of Statistics, Computer Science and Mathematics, Public University of Navarre

Disease mapping aims to study the spatial and/or spatio-temporal evolution of disease risks or rates. Three main goals are usually pursued: detect high/low risk areas, provide accurate estimates of mortality or incidence risks, and bring to light the spatio-temporal patterns. However, being statistically precise when estimating local disease risk for each area and detecting clusters with high/low-risk areas are somehow contradictory. For clustering detection tasks, methods based on scan statistics are very popular. However, these methods are not suitable for precise risk estimation at area level. By contrast, models with spatial random effects including conditional autoregressive (CAR) priors are used to smooth the risk locally by borrowing information from nearby regions (i.e. neighbor regions in space or space and time), obtaining, therefore, a more stable local risk estimates. Unfortunately, if the disease maps present clusters with high/low-risk areas, smoothing methods based on CAR priors may over-smooth local discontinuities preventing a precise estimation of disease risks.

Previous proposals introduce two-stage approaches to obtain a clustering partition of the areal units (first stage) and to estimate risks by fitting Bayesian hierarchical models including cluster effects (second stage). These proposals rely on hierarchical or density-based clustering methods for the first stage because it was shown that usual clustering methods, such as SatScan, obtained worse results, and the FleXScan method only detects high-risk clusters but not low-risk clusters. Some of these proposals include spatial and spatio-temporal dependence when detecting high-risk and low-risk clusters and when estimating disease risks. However, most of them are limited to model spatial dependence. Additionally, hierarchical and density-based clustering methods do not take into consideration the probability distribution underlying the data (as scan statistics methods do). Given the variability of the observed data, this may lead to non-optimal clustering partitions.

In this talk, we present a new approach based on scan statistics that is able to detect significant spatial and spatio-temporal high/low-risk clusters of areas. This clustering configuration is used in a second stage to improve spatial and spatio-temporal risk estimates. The behavior of the new algorithm is evaluated in a simulation study. It is subsequently used to analyze cancer mortality data in 8000 municipalities of continental Spain.

Keywords: clustering, disease mapping, risk smoothing