BAYESIAN KERNEL DENSITY ESTIMATION APPLIED TO SENSITIVE GEO-CODED DATA OF BERLIN

Ulrich Rendtel
Freie Universität Berlin, Germany, ulrich.Rendtel@fu-berlin.de

Marcus Gross
Freie Universität Berlin, Germany, Marcus.Gross@fu-berlin.de

Timo Schmid
Freie Universität Berlin, Germany, Timo.Schmid@fu-berlin.de

Sebastian Schmon
Freie Universität Berlin, Germany, Sebastian.Schmon@fu-berlin.de

Nikos Tzavidis
University Southampton, UK, N.TZAVIDIS@soton.ac.uk

Modern systems of social statistics require the timely estimation of area-specific densities of sub-populations. Ideally estimates should be based on precise geo-coded information, which is not available due to confidentiality constraints. One approach for ensuring confidentiality is by rounding the geo-coordinates. We propose multivariate non-parametric kernel density estimation that reverses the rounding process by using a Bayesian measurement error model. The methodology is applied to the Berlin register of residents for deriving density estimates of ethnic minorities and aged people. Estimates are used for identifying areas with a need for new advisory centres for migrants and infrastructure for older people.

References

Carroll, R., D. Ruppert, L. Stefanski, and C. Crainiceanu (2006). Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Scott, D. W. and S. J. Sheather (1985). Kernel density estimation with binned data. Communications in Statistics - Theory and Methods 14 (6), 1353-1359.

Wang, B. and W. Wertelecki (2013). Density estimation for data with rounding errors. Computational Statistics & Data Analysis 65, 4-12.