Density estimation over spatio-temporal data streams

Aboubacar Amiri, Sophie Dabo-Niang

PII: S2452-3062(17)30073-4
DOI: 10.1016/j.ecosta.2017.08.005
Reference: ECOSTA 80

To appear in: *Econometrics and Statistics*

Received date: 30 December 2015
Revised date: 24 August 2017
Accepted date: 26 August 2017

Please cite this article as: Aboubacar Amiri, Sophie Dabo-Niang, Density estimation over spatio-temporal data streams, *Econometrics and Statistics* (2017), doi: 10.1016/j.ecosta.2017.08.005

Density estimation over spatio-temporal data streams

Aboubacar Amiri[a], Sophie Dabo-Niang[a,b,*]

[a] *Univ. Lille, UMR 9221 - LEM - Lille Économie Management, F-59000 Lille, France.*
[b] *INRIA-MODAL Team.*

Abstract

In the last few years, data can be collected extremely easily in many scientific research fields. This became possible by the recent technological advances that have made online monitoring possible. In such situations, if real time or online estimations are expected, the usual nonparametric techniques rapidly require a lot of time to be computed and therefore become useless in practice. Adaptative counterparts of the classical kernel density estimators, that can be updated extremely easily when a new set of observations is available are investigated, for spatio-temporal processes with weak dependence structures. Mean square, uniform almost sure convergences and a central limit result are obtained under general and easily verifiable conditions. The efficiency of the considered estimators is evaluated through simulations and a real data application. The results show that the proposed method works well within the framework of a spatio-temporal data stream.

Keywords: kernel density, spatio-temporal processes, weakly dependent data, recursive kernel.
2000 MSC: 62L12, MSC 62L20, MSC 62G05, 62G07.

1. Introduction

Density estimation is a common but useful technique in statistics. It is a fundamental problem in numerical analysis, data mining and many scientific research fields, and it is also necessary for some nonparametric prediction models. There are many circumstances under which it is essential to know the density function of a specific distribution, given a sequence of random variables identically drawn from it. For instance, by knowing the density distribution of univariate or multivariate sample data, we can get an idea of the distribution of the sample. Consequently, we can calculate the mean, median and other essential quantities. For an extensive overview on the use of density estimation in statistical applications, we refer the reader to the recent work of [32], which considers nonparametric kernel-type estimation for modes that maximize nonparametric kernel-type density estimators.
In this work, we are interested in estimating a multivariate spatio-temporal random process density function. Spatio-temporal data naturally arise in many fields, such as environmental sciences, geophysics, oceanography, soil science, econometrics, epidemiology, environmental science, forestry, image processing and many others in which the phenomena of interest are continuous in space and time and the data are collected across time as well as space. A plethora of processes, such as atmospheric pollutant concentrations, precipitation fields and surface winds, are characterized by spatial and temporal variability. For some background in parametric spatial statistical modeling, refer to [40], [12], [5], [25] and the references therein. Nonparametric methods for spatial data have also been developed by many authors in past decades. For instance, kernel density estimators for spatial data have been discussed in [46] and abundantly studied in [29, 27], [6], [23], [8], [15] and [22]. Recently, [14] proposed spatial density estimators for multivariate data, depending on two kernels, one of which controls the distance between observations and the other which controls the spatial dependence structure. Additionally, [37] proposed nonparametric kernel estimators for density functions in

*Correspondence author