School of Mathematical and Statistical Sciences
Arizona State University

M E M O R A N D U M

DATE: 5/9/2024

TO: Faculty and Students

FROM: Professor(s) Sebastien Motsch
Chair/Co-Chairs of Atta Ullah
Defense for the PhD in Applied Mathematics
Committee Members

Malena Espanol
John Fricks
Nicolas Lanchier
Rodrigo Platte

DEFENSE ANNOUNCEMENT

Candidate: Atta Ullah

Defense Date: 5/20/2024

Time: 11 AM

Virtual Meeting Link: https://asu.zoom.us/j/83391341999

Tempe Campus Room: WXLR A206

Title: Voronoi tessellation and non-linear diffusion for density estimation

Please share this information with colleagues and other students, especially those studying in similar fields. Faculty and students are encouraged to attend. The defending candidate will give a 40-minute talk, after which the committee members will ask questions. There may be time for questions from those in attendance. However, guests are primarily invited to attend as observers and will be excused when the committee begins its deliberations or if the committee wishes to question the candidate privately.

ABSTRACT
-See next page-

PO Box 871804 Tempe, AZ 85287-1804
(480) 965-3951 Fax: (480) 965-8119
http://math.asu.edu

ABSTRACT

Density estimation is ubiquitous in statistical modeling and machine learning. It aims to reconstruct a probability distribution from a dataset. In this work, a non-parametric approach is developed using Voronoi tessellation and non-linear diffusion. The basic tessellation method introduces high variance in estimates. To reduce this variance, a consensus model is proposed, formulated as a system of linear ordinary differential equations, to continuously modify the dataset before applying Voronoi tessellation estimation. A regularization parameter (time) is fine-tuned by optimizing the mean integrated squared error (MISE) and least squares cross-validation (LSCV) criteria. While LSCV is less precise than MISE for selecting the optimal parameter, it has the advantage of not requiring the true distribution of the underlying data, making it more practical. One issue with regularization through consensus models is the buildup of density near the boundary. To mitigate this effect, weights are introduced into the consensus models to enforce a specific behavior of the regularizing sample at large times. Notably, using weights taken from a Gaussian distribution result in a superior fit with lower mean squared error. Finally, this approach generalized to two dimensional space. Here, the natural order of a one-dimensional space can no longer be relied on. Instead, Delaunay triangulation is used to determine the neighboring graph of the dataset. This graph allows to generalize the consensus model in higher dimensions and to define a regularization method for tessellation estimation. Numerical examples are provided to illustrate the method further.