A Mixture-based Framework for Nonparametric Density Estimation

Chew-Seng Chee

Thesis Consent Form

THE UNIVERSITY OF AUCKLAND
LIBRARY

This thesis may be consulted for the purposes of research or private study provided that due acknowledgement is made where appropriate and that permission is obtained before any material from the thesis is published. Students who do not wish their work to be available for reasons such as pending patents, copyright agreements, future publication or to protect confidential information should seek advice from the Graduate Centre as to restricted use or embargo.

Author of thesis	Chew-Seng Chee
Title of thesis	A Mixture-based Framework for Nonparametric Density Estimation
Name of degree	PhD
Date Submitted	December 2011

Print Format ((Tick the boxes that apply)	
✓	I agree that the University of Auckland Library may make a copy of this thesis available for the collection of another library on request from that library.
✓	I agree to this thesis being copied for supply to any person in accordance with the provisions of Section 56 of the Copyright Act 1994.

Digital Format - PhD theses

I certify that a digital copy of my thesis deposited with the University will be the same as the final officially approved version of my thesis. Except in the circumstances set out below, no emendation of content has occurred and I recognise that minor variations in formatting may occur as a result of the conversion to digital format.

Access to my thesis may be limited for a period of time specified by me at the time of deposit. I understand that if my thesis is available online for public access it can be used for criticism, review, news reporting, research and private study.

Digital Format– Masters theses

I certify that a digital copy of my thesis deposited with the University will be the same as the print version submitted for examination. Except in the circumstances set out below, no emendation of content has occurred and I recognise that minor variations in formatting may occur as a result of the conversion to digital format.

Access will normally only be available to authenticated members of the University of Auckland, but I may choose to allow public access under special circumstances. I understand that if my thesis is available online for public access it can be used for criticism, review, news reporting, research and private study.

Copyright (Digital Format Theses) (Tick **ONE box only**)	
✓	I confirm that my thesis does not contain material for which the copyright belongs to a third party, ***(or)*** that the amounts copied fall within the limits permitted under the Copyright Act 1994.
☐	I confirm that for all third party copyright material in my thesis, I have obtained written permission to use the material and attach copies of each permission, ***(or)*** I have removed the material from the digital copy of the thesis, fully referenced the deleted materials and, where possible, provided links to electronic sources of the material.

Signature [signature] Date 2 December 2011

Comments on access conditions

Graduate Centre only: Digital copy accepted ☐ Signature Date

A Mixture-based Framework for Nonparametric Density Estimation

Chew–Seng Chee

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy in Statistics,

The University of Auckland,

2011.

Abstract

The primary goal of this thesis is to provide a mixture-based framework for nonparametric density estimation. This framework advocates the use of a mixture model with a nonparametric mixing distribution to approximate the distribution of the data. The implementation of a mixture-based nonparametric density estimator generally requires the specification of parameters in a mixture model and the choice of the bandwidth parameter. Consequently, a nonparametric methodology consisting of both the estimation and selection steps is described. For the estimation of parameters in mixture models, we employ the minimum disparity estimation framework within which there exist several estimation approaches differing in the way smoothing is incorporated in the disparity objective function. For the selection of the bandwidth parameter, we study some popular methods such as cross-validation and information criteria-based model selection methods. Also, new algorithms are developed for the computation of the mixture-based nonparametric density estimates.

A series of studies on mixture-based nonparametric density estimators is presented, ranging from the problems of nonparametric density estimation in general to estimation under constraints. The problem of estimating symmetric densities is firstly investigated, followed by an extension in which the interest lies in estimating finite mixtures of symmetric densities. The third study utilizes the idea of double smoothing in defining the least squares criterion for mixture-based nonparametric density estimation. For these problems, numerical studies whether using both simulated and real data examples suggest that the performance of the mixture-based nonparametric density estimators is generally better than or at least competitive with that of the kernel-based nonparametric density estimators. The last but not least concern is nonparametric estimation of continuous and discrete distributions under shape constraints. Particularly, a new model called the discrete k-monotone is proposed for estimating the number of unknown species. In fact, the discrete k-monotone distribution is a mixture of specific discrete beta distributions. Empirical results indicate that the new model outperforms the commonly used nonparametric Poisson mixture model in the context of species richness estimation. Although there remain issues to be resolved, the promising results from our series of studies make the mixture-based framework a valuable tool for nonparametric density estimation.