Functional Data Analysis with Applications in Biostatistics

Sarah Jane Ratcliffe

B.Sc.(Hons), University of Technology, Sydney, 1997

Submitted in fulfilment of the requirement for the degree of

Doctor of Philosophy in the Department of Statistics,

Division of Economic and Financial Studies, Macquarie University.

October, 2000

Contents

Summary

Functional data analysis is concerned with the analysis of data for which the observed responses for each subject are continuous curves. In practice, measurements are taken at discrete time points but estimates are required over the entire time interval. Traditional techniques for analysis of multiple curves, such as longitudinal data analysis or time series methods, are unsuitable for this type of data, since there are generally more measurements per subject than subjects and stationarity assumptions do not necessarily hold. With a technology induced growth in data of this kind, research into techniques for functional data analysis has become an emerging area in recent years.

This thesis aims to develop new techniques for functional data analysis, focusing on three problems: logistic regression with a functional regressor, linear and logistic regression for a repeatedly stimulated functional regressor, and a functional mixed-effects type model for joint mean and covariance modelling.

For each of the problems, we develop solutions using a basis function approach, that is, expressing the data for each subject as a linear combination of known basis functions. Using this approach we are able to overcome singularity problems associated with having more measurements than subjects. As well as calculating maximum likelihood or least squares parameter estimates, model diagnostic and smoothing parameter selection issues are addressed.

The techniques developed in this thesis are applied to novel biostatistical data sets: electroencephalographic data and fetal heart rate data. Of main interest is the fetal heart rate data, which motivated the development of the regression techniques for a repeatedly stimulated

functional parameter. It was found that the stimulated fetal heart rates could be used to predict an infant's risk category at birth and psychomotor development at 18 months of age.

Most of the material presented in the thesis is my own work. The exception is:

1. the work described in Section 6.3 is partly due to Victor Solo.

Acknowledgements

I would like to take this opportunity to thank a number of people, without whom this thesis would not have been possible.

Firstly, I wish to thank my supervisors, Professor Victor Solo and Dr Gillian Heller, for all of their help, encouragement and guidance over a number of years. The time spent working as a student under their supervision has been an invaluable learning experience.

Thanks also to Dr Leo Leader, from the School of Obstetrics and Gynaecology, University of New South Wales and the Royal Hospital for Women, for providing the fetal data, feedback on the analyses, medical background and other useful inputs. It has been great working with you throughout this thesis.

Dr Evian Gordon from the Department of Psychiatry, University of Sydney and Westmead Hospital provided the EEG data, while Dr Julian Leslie arranged for its use in this thesis. Thank you.

I am indebted to members of the Statistics Department of Macquarie University for their advice and encouragement throughout my study period. Also, thanks to many friends for their continued moral support.

Finally, I thank my parents, Fred and Beverley, and brother James for all of their support and patience during this degree.

Certificate

The work described in this thesis was carried out in the Department of Statistics, Division of Economic and Financial Studies, Macquarie University, New South Wales, Australia, between March 1997 and October 2000, under the supervision of Professor Victor Solo and Dr Gillian Heller.

This is to certify that the material presented in this thesis is original and, to the best of my knowledge and belief, contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma at a university or other institute of higher learning, except where due acknowledgement is made in the text.

October, 2000.

Sarah Jane Ratcliffe

List of Figures

List of Tables