INFERENCE FOR DISCRETE TIME STOCHASTIC PROCESSES USING AGGREGATED SURVEY DATA

A thesis submitted for the degree of
Doctor of Philosophy
of The Australian National University

Brett Andrew Davis
February 2003

DECLARATION OF AUTHENTICITY

This thesis is my own work and all sources used have been acknowledged.

Brett Andrew Davis

February 2003

ACKNOWLEDGEMENTS

I am very grateful to my supervisors, Emeritus Professor C. R. Heathcote and Professor T. J. O'Neill, for their guidance and encouragement throughout my course of study and in the preparation of this thesis.

My sincere thanks also go to Dr Michael Martin, Dr Borek Puza and Dr Steven Stern for their kind assistance with my study.

Special thanks are due to the following people for providing the datasets analysed in Chapters Three and Four: Dr N. Brouard (INED, Paris), Dr M. Nurminen (Finnish Institute of Occupational Health) and Dr C. Mathers (Australian Institute of Health and Welfare).

CONTENTS

SUMMARY

We consider a longitudinal system in which transitions between the states are governed by a discrete time finite state space stochastic process X. Our aim, using aggregated sample survey data of the form typically collected by official statistical agencies, is to undertake model based inference for the underlying process X. We will develop inferential techniques for continuing sample surveys of two distinct types. First, longitudinal surveys in which the same individuals are sampled in each cycle of the survey. Second, cross-sectional surveys which sample the same population in successive cycles but with no attempt to track particular individuals from one cycle to the next. Some of the basic results have appeared in Davis et al (2001) and Davis et al (2002).

Longitudinal surveys provide data in the form of transition frequencies between the states of X. In Chapter Two we develop a method for modelling and estimating the one-step transition probabilities in the case where X is a non-homogeneous Markov chain and transition frequencies are observed at unit time intervals. However, due to their expense, longitudinal surveys are typically conducted at widely, and sometimes irregularly, spaced time points. That is, the observable frequencies pertain to multi-step transitions. Continuing to assume the Markov property for X, in Chapter Three, we show that these multi-step transition frequencies can be stochastically interpolated to provide accurate estimates of the one-step transition probabilities of the underlying process. These estimates for a unit time increment can be used to calculate estimates of expected future occupation time, conditional on an individual's state at initial point of observation, in the different states of X.

For reasons of cost, most statistical collections run by official agencies are cross-sectional sample surveys. The data observed from an on-going survey of this type are marginal frequencies in the states of X at a sequence of time points. In Chapter Four we develop a model based technique for estimating the marginal probabilities of X using data of this form. Note that, in contrast to the longitudinal case, the Markov assumption does not simplify inference based on marginal frequencies. The marginal probability estimates enable estimation of future occupation times (in each of the states of X) for an individual of unspecified initial state. However, in the applications of the technique that we discuss (see Sections 4.4 and 4.5) the estimated occupation times will be conditional on both gender and initial age of individuals.

The longitudinal data envisaged in Chapter Two is that obtained from the surveillance of the same sample in each cycle of an on-going survey. In practice, to preserve data quality it is necessary to control respondent burden using sample rotation. This is usually achieved using a mechanism known as *rotation group sampling*. In Chapter Five we consider the particular form of rotation group sampling used by the Australian Bureau of Statistics in their Monthly Labour Force Survey (from which official estimates of labour force participation rates are produced). We show that our approach to estimating the one-step transition probabilities of X from transition frequencies observed at incremental time intervals, developed in Chapter Two, can be modified to deal with data collected under this sample rotation scheme. Furthermore, we show that valid inference is possible even when the Markov property does not hold for the underlying process.