ESRC 'e-STAT' PhD Studentship

Using simulation and parallel computing for faster sample size calculations in complex random effect models

Social Statistics PhD

Centre for Multilevel Modelling, University of Bristol

This PhD will form part of an ESRC-funded project, 'e-STAT' funded as a new Node under the National Centre for E-Social Science. The 'e-STAT' project is developing generic statistical multilevel modelling methodology for complex modelling across a range of applications in the social sciences. Estimation algorithms for the models are being developed using MCMC techniques.

A brief description of the project is as follows:

It is well known that the dependence induced by clustering in social science datasets means that the sample size requirements for testing hypotheses need inflating to account for the lack of independence. In a recently completed ESRC funded project a piece of software, MLPOWSIM has been developed that will generate both MLwiN macro code and R code to perform sample size calculations for a selection of multilevel nested and crossed designs. The software is limited to 2 separate clustering factors (whether they be nested or crossed) and it is challenging to consider how to simulate realistic datasets for specific crossed scenarios.

For example in education a study of the effects of schools and neighbourhoods on achievement would typically consider sampling groups of pupils within a selection of schools with the neighbourhood identifiers being collected but not part of the sampling design. Most sample size calculations assume a balanced sampling design but it will clearly not be possible to sample exactly 10 pupils from each school and neighbourhood combination as some schools will not teach pupils from some neighbourhoods and so using simulation to generate samples from a two way table of pupil counts within schools by neighbourhoods is a potential solution.

This has been done in MLPOWSIM for the case of 2 clustering factors but in reality datasets exist with many classifications. For example in recent work Leckie (2008) examined using the National pupil database to look at accounting for the effects of primary school, secondary school, neighbourhood on student achievement while also controlling for student mobility. This database contains nearly half a million pupil records which makes using the whole database rather unwieldy. There is therefore a need to come up with simulation-based approaches for generating appropriate samples from such a large database to establish potential sample sizes and sampling schemes that will capture the basic structure of the data and have appropriate power for testing hypotheses both using the current data and future years of data. This challenge is the motivation for the PhD project.

One also needs to investigate how to come up with parameter values to be used in the simulated designs and how sensitive the sample size estimates are to these values. Another factor in increasing the number of classifications is that the estimation options for the models become more limited, and in particular Monte Carlo Markov chain (MCMC) estimation becomes more commonly the method used. MCMC estimation is inherently slow as it can involve running for many iterations of the algorithm. When combined with the simulation approach of producing thousands of datasets the time to construct a sample size calculation becomes very large. Here parallel computing can very easily cut the time as it is possible to send different datasets to different processors and link the results together at the end. It is also possible that parallel computing may well speed up individual runs of the MCMC algorithm by parallelizing steps of the MCMC algorithm.

For UK nationals, the studentship will attract an annual stipend of **£12,600** in addition to tuition fees

The Centre for Multilevel Modelling is a multidisciplinary cross-department research centre based mainly in the Graduate School of Education (GSOE) but with researchers in both Geography and Clinical Veterinary Science. The internationally-renowned multilevel modelling research team are involved in advancing statistical methods for analysing large-scale datasets and investigate issues such as school effectiveness, the effect of school resources on pupil attainment and family effects on child development. The Centre also produces the software 'MLwiN'.

The student would be jointly supervised by Professor William Browne and Professor Jon Rasbash. Professor Browne is based in the Department of Clinical Veterinary Science whilst Professor Rasbash is the director of the Centre for Multilevel Modelling and is based in the GSOE. The student will formally be enrolled on the PhD within the GSOE but will physically be based at the vet school in Lower Langford with Professor Browne and his researchers.

The team at the Centre for multilevel modelling includes applied statisticians, theoretical statisticians, software engineers, computational statisticians, psychologists, geographers and educationalists. We provide a lively research environment and can offer support for applicants wishing to develop statistical methodology skills and applied statistical modelling skills. The team has considerable expertise in statistical programming and software engineering so additional support can be given to candidates wishing to improve skills in these further two areas.

Applicants should have a masters with a high content of applied statistics including

- methods of research design and data collection
- methods of data analysis
- a critical perspective on how statistical methods are used to address substantive research issues in the social sciences

For more information of these criteria see section F18 subsection 4.1.2 of the ESRC Postgraduate Training Guidelines

http://www.esrc.ac.uk/ESRCInfoCentre/Images/Postgraduate_Training_Guidelines_2005_tcm6-9062.pdf

Bristol is a vibrant city with a large student population and excellent transport links.

Applicants: It is anticipated that the studentship will begin as soon as possible and the latest in January 2010.

For further information in advance of submitting an application, email Professor Bill Browne (william.browne@bristol.ac.uk) or Professor Jon Rasbash (j.rasbash@bristol.ac.uk).

To apply for this studentship online go to
http://www.bris.ac.uk/prospectus/postgraduate/2010/intro/8

(Please note your interest in this studentship in the 'Funding' section of the application form by using the reference NCeSS).

For information about the application process, contact ed-phd@Bristol.ac.uk, using the reference NCeSS.

The closing date is 6th November 2009.

http://www.cmm.bristol.ac.uk/   http://www.bristol.ac.uk/education/