Final Report on SCB Program.

The organizers were Profs. Steve Brooks (Cambridge), Arnaud Doucet (UBC), Paul Fearnhead (Lancaster), Wally Gilks (Leeds) and Simon Tavaré (USC/Cambridge).

Our theme of Stochastic Computation in the Biological Sciences took place at the Newton Institute between October 23 and December 15 2006. We organized three one-week workshops during the program: (i) Recent Advances in Monte Carlo Based Inference; (ii) Stochastic Computation for the Analysis of Ecological and Epidemiological Data and (iii) Recent Advances in Statistical Genetics and Bioinformatics.

The first workshop, organized by Doucet and Fearnhead and running between 30 October and 3 November, was devoted to advances in methdology in stochastic computation. Monte Carlo methods, particularly Markov Chain Monte Carlo (MCMC), are often the methods of choice for making inferences about complex stochastic systems. Whilst MCMC dates back over 50 years, and there has been extensive research in Monte Carlo methods over the past 20 years, there are still many challenges that face researchers today.
These include analysis of highly complicated stochastic models and large scientific data sets and understanding the theoretical properties of some of the recent novel approaches.

Currently, Monte Carlo methods are used by researchers in numerous scientific fields, including statistics, physics, engineering, genetics, econometrics, bioinformatics, and machine learning. This interdisciplinary workshop brought together researchers from a variety of such fields to discuss current and novel Monte Carlo methodology, and to cross-fertilise ideas across these different disciplines. The workshop had a broad focus, covering both recent advances in more established methods such as MCMC and sequential Monte Carlo, together with more recent ideas that have had little exposure within the statistics community, such as Variational Methods, Population Monte Carlo, Approximate Bayesian Computation, Quasi Monte Carlo, and Indirect Inference.

The second workshop, organized by Brooks and running between 20 November and 24 November, focussed on statistical issues arising in epidemiology and ecology. Both areas pose similar methodological problems for statisticians. Both require the analysis of large yet sparse datasets often with both individual and environmental time-varying explanatory factors. In both cases effective management is the key goal as well as developing understanding as to the processes underpinning the observed dynamics.
The workshop explored the cutting-edge of scientific developments in these fields and discussed the statistical methodologies required to underpin future advances. In particular, the computational and statistical challenges involved in modelling animal population and disease dynamics at the individual level were discussed, and experts from a range of disciplines helped to outline future scientific and statistical developments.

The final workshop, organized by Gilks and Tavaré and running between December

11 and 15, focussed on topics from bioinformatics and statistical genetics. Bioinformatics has grown up around the rapidly expanding databases of DNA sequence, molecular structure and interaction. Understanding the evolutionary and functional relationships between these biomolecules is perhaps the greatest challenge of 21st century science. Advanced statistical methods are becoming increasingly important in this endeavour. The arrival of dense genotyping arrays has provided huge amounts of SNP data which have been used successfully in recentin efforts to map genes associated with complex traits. Statistical methods for designing and analyzing such whole genome scans are beginning to appear, but many open problems remain. The use of SNP data to estimate recombination rates across the genome, to detect evidence of selective sweeps and to infer demographic history have also provided statistical challenges that were addressed in the workshop.

Professor Elizabeth Thompson (University of Washington) was invited as the Rothschild Visiting Professor. She presented a special lecture on Monday 11 December entitled 'Uncertainty and evidence in likelihoods for genetic linkage.' The theme of the talk was the notion of fuzzy p-values, a topic the audience found most stimulating and provocative.

In addition to the workshops there was a very active group of resident scientists and mathematicians, numbering some 40 people over the course of the program, who contributed to the lively environment in the Newton Institute. Our interactions developed around weekly informal 'organizational' meetings in which topics of interest were discussed. A number of ongoing collaborations flourished, and several new ones developed from these interactions, including those on efficient posterior simulation of large decomposable graphs (Peter Green and Alun Thomas), on statistical inference in the fossil record (Simon Tavaré and Bruce Rannala), and on the analysis copy number variation (Simon Myers and Matt Hurles), to name just three.

Many of the visitors commented on the fact that the INI provides a great environment in which to focus on research without the distractions of regular university life. The organizers express their real appreciation for the support provided at all levels by the INI staff. A truly wonderful place to visit, and wonderful people to interact with!