Mathematics and Statistics Undergraduate Handbook
Supplement to the Handbook

Honour School of Mathematics and Statistics Syllabus and Synopses for Part C 2006–07 for examination in 2007

October 2006

Contents

Every effort is made to ensure that the list of courses offered is accurate at the time of going online (Trinity Term 2006). However, students are advised to check the up-to-date version of this document on the Department of Statistics website in September 2006.
Notice of misprints or errors of any kind, and suggestions for improvements in this booklet should be addressed to the Academic Administrator in the Department of Statistics.

1 Honour School of Mathematics and Statistics

[See the current edition of the Examination Regulations for the full regulations governing these examinations.]

Students staying on to take the four-year course will take 2 units from Part C in their fourth year, and will also offer a dissertation on a statistics project. Of the 2 units from Part C, at least half a unit will be from the schedule of ‘Statistics’ units for Part C. This booklet describes the units and half-units available in Part C. Information about dissertations/ statistics projects is available on the Department of Statistics website.

In the classification awarded at the end of the degree, unit paper marks in Part A will be given a ‘weighting’ of 2, and unit marks in Part B will be given a ‘weighting’ of 3. For those students staying on to do the fourth year, the Part C unit marks will also be taken into account with a ‘weighting’ of 4.

We ask that you register by the end of week 9 Trinity Term 2006 for classes for the Mathematics/ Statistics courses that you wish to take. A registration form is attached to these synopses. Some combinations of subjects are not advised and lectures in these subjects may clash. However, when timetabling lectures we will aim to keep clashes to a minimum.

1.1 Units and half-units and methods of examination

Most of the lecture courses given are available as half-units. Half units are examined in an examination paper of 1 ¾ hours, unless the synopsis says otherwise; whole units will be examined in a 3 hour exam unless the synopsis states otherwise.

All the units and half-units described in this booklet are “M-level”.

2 Statistics units and half units

2.1 **MS1a: Graphical Models and Inference** — Professor S Lauritzen- 16MT

Recommended Prerequisites
BS1 Applied statistics and BS2 Statistical Inference would be helpful but not essential.

Aims & Objectives
Graphical models have become increasingly important in many areas where statistics play a role. They enable the description and analysis of complex stochastic systems via their natural modularity, expressed in terms of (mathematical) graphs which encode conditional independence structure. The modules correspond typically to well-understood, classical models. This course builds upon develops the specific theory and computational tools needed in the analysis of graphical models for categorical and multivariate Gaussian data as well as Bayesian graphical models for complex stochastic systems.

Synopsis
Topics in MT06 include:
1. Conditional independence and Markov properties.
2. Log-linear graphical models for categorical data.
3. Gaussian graphical models.
4. Graphical models for complex stochastic systems

Method of Examination
4 examination questions, 1 ¾ hour paper

Reading
1. D. Edwards, Introduction to Graphical Models (2nd ed.), Springer-Verlag, New York (2002).
2. S. L. Lauritzen, Graphical Models, Oxford University Press, Oxford (1996).
3. P. J. Green, N. L. Hjort and S. Richardson, eds. Highly Structured Stochastic Systems, Oxford University Press, Oxford (2003).

2.2 **MS1b: Statistical Data Mining** — Dr G Nicholls- 12HT plus 4 1-hour computer practical classes

Recommended Prerequisites
Part A Probability and Statistics. BS1 Applied Statistics would be an advantage.

Aims & Objectives
'Data mining' is now widely used to find interesting patterns in large databases, for example in insurance, in marketing and in many scientific fields. With large amounts of data we can search for quite subtle patterns.

This course concentrates on the statistical tools used to identify patterns, and then to identify those which are interesting not just the result of chance associations.

Synopsis
Fundamentals of pattern recognition, machine learning and data mining.

Exploratory methods: principal components analysis, biplots, independent component analysis, multidimensional scaling.

Cluster Analysis: K-means, hierarchical methods, vector quantisation, self-organising maps.

Linear discriminant analysis, logistic discrimination, linear separation and perceptrons.

Classification trees. Splitting criteria, existence of pruning sequences. V-fold cross-validation.

Feed-forward neural networks. Universal approximation properties, back-propagation, training algorithms, assessment of fit.

Method of Examination
4 examination questions

Reading
1. C. Bishop, Neural Networks for Pattern Recognition, Oxford UP (1995).
2. D. Hand, H. Mannila, P. Smyth, Principles of Data Mining, MIT Press (2001).
3. I. H. Witten and E. Franke, Data Mining. Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann (2000).

Further Reading
1. B. D. Ripley, Pattern Recognition and Neural Networks, Cambridge UP (1996).

2.3 **MS2a: Bioinformatics and Computational Biology** — Dr G Nicholls and Professor J Hein - 16MT

Recommended Prerequisites
None. In particular, no previous knowledge of Genetics will be necessary.

Aims & Objectives
Modern molecular biology generates large amounts of data, such as sequences, structures and expression data, that needs different forms of statistical analysis and modelling to be properly interpreted. The fields of Bioinformatics and Computational Biology have this as their subject matter and there is no sharp boundary between them. Bioinformatics has an applied flavour while Computational Biology is viewed as the study of the models, statistical methodology and algorithms needed to do bioinformatics analysis. This course aims to present core topics of these fields with and emphasis on modelling and computation.

Synopsis
Fundamental Data Structures in Biology: Sequences, Genes and RNA secondary structure. Stochastic Models of Sequence and Genome Evolution including models of single nucleotide/amino acid/codon evolution.

Phylogenies: enumerating phylogenies, the probability of sequences related by a specified phylogeny, the minimal number of events needed to explain a data set (Parsimony).

Likelihood and algorithms (Markov Chain Monte Carlo) for inference based on the likelihood. Software packages for sample-based inference.

Alignment Algorithms. Comparing 2 strings, an arbitrary number of strings, find segments of high similarity in 2 strings.

Common Patterns in a set of Sequences.

Method of Examination
4 examination questions.

Reading

1. C. Semple and M. Steel, Phylogenetics, Oxford University Press (2003).
2. Durbin et al., Biological Sequence Analysis, Cambrigde University Press (1998).
3. T. Jiang et al., (editors) Current Topics in Computational Biology, MIT Press, (2003).
4. M. S. Waterman et al., Computational Genome Analysis: An Introduction, Springer (2004).

2.4 **MS2b: Stochastic Models in Mathematical Genetics** — Prof B Griffiths and Prof A Etheridge — 16HT

Aims & Objectives
The aim of the lectures is to introduce modern Stochastic models in Mathematical Population Genetics that describe the distribution of gene frequencies and ancestry in a population or sample of genes. Stochastic and Graph theoretic properties of coalescent and gene trees are studied in the first eight lectures. Diffusion process models of gene frequencies and their applications are studied in the second eight lectures.

Synopsis
Evolutionary models in Mathematical Genetics:
The Wright-Fisher model. The Genealogical Markov chain describing the number ancestors back in time of a collection of genes.

The Coalescent process describing the stochastic behaviour of the ancestral tree of a collection of genes. Mutations on ancestral lineages in a coalescent tree. Inferring the time to the most recent common ancestor in a sample of genes from the number of mutations occurring to the genes. Models with a variable population size.

The frequency spectrum and age of a mutation. Ewens' sampling formula for the probability distribution of the allele configuration of genes in a sample in the infinitely-many-alleles model. Hoppe's urn model for the infinitely-many-alleles model.

The infinitely-many-sites model of mutations on DNA sequences. Gene trees as perfect phylogenies describing the mutation history of a sample of DNA sequences. Graph theoretic constructions and characterizations of gene trees from DNA sequence variation. Gusfield's construction algorithm of a tree from DNA sequences. Examples of gene trees from data. The probability distribution of a gene tree.

Diffusion process models in Mathematical Genetics:
Introduction to diffusion processes. The stochastic process describing the distribution of the gene frequency of an allele forward in time for a two-allele model. The Moran model. The diffusion process limit from the Moran model. The generator of a diffusion process with two allele types.

Heuristic introduction to Stochastic differential equations. Examples in using the diffusion process generator and Stochastic differential equations. The mean time to absorption or fixation of an allele.

The genealogy of the diffusion process describing the gene frequency of an allele. The underlying infinite-particle coalescent.

Two allele models with mutation and selection. Stationary distributions of diffusion process. Sampling from the stationary distribution.

A brief introduction to diffusion process models with more than two types. The Dirichlet distribution describing the stationary distribution of allele frequencies. The Poisson-Dirichlet process.

Method of Examination
4 examination questions.

Reading
1. R. Durrett, Probability Models for DNA Sequence Evolution, Springer (2002).
2. W. J. Ewens, Mathematical Population Genetics, 2nd ed, Springer (2004).
3. J. R. Norris, Markov Chains, Cambridge University Press (1999).
4. M. Slatkin and M. Veuille, Modern Developments in Theoretical Population Genetics, Oxford Biology (2002).
5. S. Tavar´e and O. Zeitouni, Lectures on Probability Theory and Statistics, Ecole d'Et´e de Probabilit´es de Saint-Flour XXXI - 2001, Lecture Notes in Mathematics 1837. Springer (2004).

2.5 **MS3b: Lévy Processes and Finance** — Dr Winkel — 16HT

Method of examination
4 examination questions.

Prerequisites
Part A Probability is a prerequisite. BS3a/OBS3a Applied Probability or B10 Martingales and Financial Mathematics would be useful, but are by no means essential; some material from these courses will be reviewed without proof.

Aims
Lévy processes form a central class of stochastic processes, contain both Brownian motion and the Poisson process, and are prototypes of Markov processes and semimartingales. Like Brownian motion, they are used in a multitude of applications ranging from biology and physics to insurance and finance. Like the Poisson process, they allow to model abrupt moves by jumps, which is an important feature for many applications. In the last ten years Lévy processes have seen a hugely increased attention as is reflected on the academic side by a number of excellent graduate texts and on the industrial side realising that they provide versatile stochastic models of financial markets. This continues to stimulate further research in both theoretical and applied directions. This course will give a solid introduction to some of the theory of Lévy processes as needed for financial and other applications.

Synopsis
Review of (compound) Poisson processes, Brownian motion (informal), Markov property.

Connection with random walks, [Donsker's theorem], Poisson limit theorem. Spatial Poisson processes, construction of Lévy processes.

Special cases of increasing Lévy processes (subordinators) and processes with only positive jumps. Subordination. Examples and applications. Financial models driven by Lévy processes. Stochastic volatility. Level passage problems. Applications: option pricing, insurance ruin, dams.

Simulation: via increments, via simulation of jumps, via subordination. Applications: option pricing, branching processes.

Reading
1. J.F.C. Kingman: Poisson processes, Oxford University Press (1993), Ch.1-5, 8.
2. A.E. Kyprianou: Introductory lectures on fluctuations of Lévy processes with Applications, Springer (2006), Ch. 1-3, 8-9.
3. W. Schoutens: Lévy processes in finance: pricing financial derivatives, Wiley (2003).

Further reading
1. J. Bertoin: Lévy processes, Cambridge University Press (1996), Sect. 0.1-0.6, I.1, III.1-2, VII.1.
2. K. Sato: Lévy processes and infnite divisibility,Cambridge University Press (1999), Ch. 1-2, 4, 6, 9.

2.6 MS4b/C11.1b: **Probabilistic Combinatorics** — Dr Martin — 16HT

[In the synopses booklet for Mathematics Part C, this course appears in the Mathematics Deparment Units section; in this booklet it is in the Statistics section. For any Mathematics and Statistics student taking this half-unit, it will count as a Statistics half-unit. Note that a recommended prerequisite is C11.1a Graph Theory, which is available to Mathematics and Statistics students in Section 3 – C11.1a counts as a Mathematics half-unit for Mathematics and Statistics students.]

Recommended Prerequisites
C11.1a Graph Theory. Probability up to the level of B10a is also desirable.

Learning outcomes
To develop an appreciation of probabilistic methods in discrete mathematics.

Aims and objectives
Probabilistic combinatorics is a very active field of mathematics, with connections to other areas such as computer science and statistical physics. Probabilistic methods are essential for the study of random discrete structures and for the analysis of algorithms, but they can also provide a powerful and beautiful approach for answering deterministic questions. The aim of this course is to introduce some fundamental probabilistic tools and present a few applications.

Synopsis
Spaces of random graphs. Threshold functions.

First and second moment methods. Chernoff bounds. Applications to Ramsey numbers and random graphs.

Lovasz Local Lemma. Property B.

Poisson approximation, and application to the distribution of small subgraphs. Janson's inequality.

Concentration of measure. Martingales and the Azuma-Hoeffding inequality.

Chromatic number of random graphs.

Talagrand's inequality.

Method of Examination
4 examination questions.

Reading

1. N. Alon and J.H. Spencer. The Probabilistic Method, second edition, Wiley, 2000.
Further reading:
1. B. Bollobas, Random Graphs, second edition, CUP, 2001.
2. S.Janson, T. Luczak and A.Rucinski, Random Graphs, John Wiley and Sons, 2000.
3. M. Mitzenmacher and E. Upfal. Probability and Computing : Randomized Algorithms and Probabilistic Analysis, Cambridge University Press, New York (NY), 2005.

3 Mathematics units and half units

The Mathematics units and half units that students may take are drawn from Part C of the Honour School of Mathematics. For full details of these units and half-units, see the Syllabus and Synopses for Part C of the Honour School of Mathematics, which are available on the web at http://www.maths.ox.ac.uk/current-students/undergraduates/handbooks-synopses/maths.shtml

The Mathematics units and half-units that are available are as follows:
C1.1 (or C1.1a or C1.1b) Model Theory and Gödel's Incompleteness Theorems
C1.2 (or C1.2a or C1.2b) Axiomatic Set Theory and Analytic Topology
C2.1 (or C2.1a or C2.1b) Lie Algebras and Representation Theory of Symmetric Groups
C3.1 (or C3.1a or C3.1b) Topology and Groups and Algebraic Topology
C4.1 (or C4.1a or C4.1b) Functional Analysis and Banach and C* Algebras
C4.2a Real and Harmonic Analysis
C5.1: (or C5.1a) Partial Differential Equations for Pure and Applied Mathematicians and Calculus of Variations
C6.1 Solid Mechanics
C6.2 Elasticity and Plasticity
C6.3 (or C6.3a or C6.3b) Perturbation Methods and Applied Complex Variables
C7 (or C7.1b or C7.2b) Mathematical Physics
C7.3 Advanced Quantum Mechanics
C8.1 (or C8.1a or C8.1b) Mathematics and the Environment and Mathematical Physiology
C9.1 (or C9.1a or C9.1b) Analytic Number Theory and Elliptic Curves
C10.1 (or C10.1a or C10.1b) Stochastic Differential Equations and Brownian Motion in Complex Analysis
C11.1a Graph Theory
C12.1 (or C12.1a or C12.1b) Numerical Linear Algebra and Continuous Optimization
C12.2 (or C12.2a or C12.2b) Approximation Theory and Finite Element Methods

4 Registration

We ask that students register in advance for the classes they wish to take, by the end of week 9 Trinity Term 2006, using the form overleaf.

Because of the large number of options which are available in Part C, some lectures will clash. See the Syllabus and Synopses for Part C of the Honour School of Mathematics for information on which lectures may clash.

FHS MATHEMATICS AND STATISTICS
REGISTRATION FORM: PART C CLASSES 2006–07

SURNAME . . . . .FIRST NAME . . . .
EMAIL ADDRESS . . . .
COLLEGE . . . .

Note: As described in Section 1, you need to do a total of 2 units in Part C (in addition to doing a dissertation on a statistics project).

Please give details of the subjects in which you wish to take classes.

I wish to take classes in the following subjects: [Please Tick]

_ MS1a Graphical Models and Inference

_ MS1b Statistical Data Mining

_ MS2a Bioinformatics and Computational Biology

_ MS2b Stochastic Models in Mathematical Genetics

_ MS3b Lévy Processes and Finance

_ MS4b Probabilistic Combinatorics

For Mathematics units or half-units, please list the unit or half-unit code and name:
Unit code Unit name

. . . .

. . ..

. . . .

. . . .
Please return this form to the Academic Administrator, Department of Statistics, by the end of week 9 Trinity Term 2006.