

# Mathematics and Statistics Undergraduate Handbook

## Supplement to the Handbook Honour School of Mathematics and Statistics Syllabus and Synopses for Part C 2008–2009 for examination in 2009

### Contents

0.	Changes for 2008-2009	2
1	Honour School of Mathematics and Statistics	3
1.1	Units and half-units and methods of examination	3
2	Statistics units and half units	6
2.2	MS1b: Statistical Data Mining - 12HT	6
2.3	MS2a: Bioinformatics and Computational Biology - 16MT	7
2.4	MS2b: Stochastic Models in Mathematical Genetics - 16HT	8
2.5	MS3b: Lévy Processes and Finance - 16HT	9
2.6	MS4b/C11.1b: Probabilistic Combinatorics - 16HT	10
3	Mathematics units and half units	12
4	Registration	12

*Every effort is made to ensure that the list of courses offered is accurate at the time of going online. However, students are advised to check the up-to-date version of this document on the Department of Statistics website.*

*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.*

Updated Oct 2008

## **0. Changes for 2008-2009**

**A brief summary of the changes for 2008-2009 (for examination in 2009) in comparison with 2007-2008 (examination in 2008) is as follows.**

**(1) In the examination on each half-unit there will be 3 questions set (rather than 4 as in 2007-2008).**

**(2) The rubric for 3-hour exam papers has changed: now a candidate's best two answers on each half-unit will count for the total mark.**

**(3) A half unit paper lasts 1 ½ hours (rather than 1 ¾ hours as in 2007-2008).**

**(4) We will not offer MS1a Graphical Models and Inference in 2008-2009.**

**(5) The method of assessment for MS1b *Statistical Data Mining*, and MS2a *Bioinformatics and Computational Biology*, will be by mini-project (with the project available in week 8 of the appropriate term, and due for submission at the Examination Schools by 12 noon on Monday of week 1 of the following term).**

**Similar changes for 2008-2009 (for examination in 2009) are being made to Part B Mathematics. Note that changes to Mathematics courses are not listed above.**

**Changes (1) and (2) are to ensure that candidates have attained a broad knowledge and understanding of each course at Part C (a similar change is being made to Part B).**

**When details regarding (4) and (5) are firm, the web version of this document will be updated.**

## 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 at [http://www.stats.ox.ac.uk/current\\_students/bammath/projects](http://www.stats.ox.ac.uk/current_students/bammath/projects)

We ask that you register by the end of week 9 Trinity Term 2008 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.

*Language Classes:* Mathematics and Statistics students are also invited to apply to take classes in a foreign language. In 2008-2009 classes are offered in French. Students' performance in these classes will not contribute to the degree classification in Mathematics and Statistics. However successful completion of the course may be recorded on student transcripts. See [http://www.maths.ox.ac.uk/current\\_students/undergraduates/handbooks-synopses](http://www.maths.ox.ac.uk/current_students/undergraduates/handbooks-synopses) for further information.

Candidates leaving after four years who satisfy the Examiners will be awarded an MMath in Mathematics and Statistics with two associated classifications.

For the MMath in Mathematics and Statistics in TT 2009 one of the classifications will be based on Part C alone. Each candidate will receive a numerical mark on each paper in each Part of the examination in the University standardised range 0-100, such that

- a First Class performance (on that paper) is indicated by a mark of 70 to 100;
- an Upper Second Class performance (on that paper) is indicated by a mark of 60 to 69;
- a Lower Second Class performance (on that paper) is indicated by a mark of 50 to 59
- a Third Class performance (on that paper) is indicated by a mark of 40 to 49;
- a Pass performance (on that paper) is indicated by a mark of 30 to 39;
- a performance at the level of a Fail (on that paper) is indicated by a mark of 0 to 29.

### 1.1 Units and half-units and methods of examination

The Statistics lecture courses given are available as half-units apart from MS4b/C11.b *Graph Theory/ Probabilistic Combinatorics*. Half units are examined in an examination paper of 1 ½ hours or by mini-project.

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

### **The Examination Papers in Part C**

Each 16 hour lecture course is assessed as a half unit. On each half unit examination paper of 1 ½ hours, three questions will be set. Only the two best questions on that half unit paper count.

Most Mathematics papers including MS4b/C11.b *Graph Theory / Probabilistic Combinatorics* are whole unit papers, that is, they contain the questions for two half units. There are three questions set for each half unit. If you are taking a full unit, then you may attempt as many questions as you like but only your two best answers on each half unit will count towards your final mark for the paper.

Mini-project assignments can be collected from the Department of Statistics in week 8 of the relevant term and should be submitted to the Examination Schools by noon Monday week 1 of the following term.

### **Marking of Papers in Part C**

For the Statistics Department papers, each question will initially be marked out of 25 and marking schemes for the questions will aim to ensure that the following qualitative criteria hold:

20-25 marks. A completely, or almost completely, correct answer, showing excellent understanding of the concepts and skill in carrying through the arguments and/or calculations.

13-19 marks. A good though not complete answer, showing understanding of the concepts and competence in handling the arguments and/or calculations. Such an answer might consist of an excellent answer to a substantial part of the question, or a good answer to the whole question which nevertheless shows some flaws in calculation or in understanding or in both.

### **Analysis of Marks in Part C**

The Board of Examiners in Part C will assign USMs for full unit and half unit papers taken in Part C and may recalibrate the raw marks to arrive at university standardised marks reported to candidates. The full unit papers are designed so that the raw marks sum to 100, however, Examiners will take into account the relative difficulty of papers when assigning USMs. In order to achieve this, Examiners may use information on candidates' performances on the earlier Parts of the examination when recalibrating the raw marks. They may also use other statistics to check that the USMs assigned fairly reflect the students' performances on a paper.

The USMs awarded to a candidate for papers in Part C will be used to arrive at a classification for Part C of the MMath.

- First class  $70 \leq Av \text{ USM Part C}$
- Upper Second Class  $60 \leq Av \text{ USM Part C} < 70$
- Lower Second Class  $50 \leq Av \text{ USM Part C} < 60$

- Third class  $40 \leq \text{Av USM Part C} < 50$

[Note: Half unit papers count as half a paper when determining the average USM.]  
Candidates leaving after four years who satisfy the Examiners will be awarded an MMath in Mathematics and Statistics, with two associated classifications; for example: MMath in Mathematics and Statistics: Years 2 and 3 together - First Class; Year 4 – First Class

A pass degree will not be awarded for Year 4. Candidates achieving

$\text{AvUSM Part C} < 40$  may supplicate for a BA.

### **Class Descriptors**

The average USM ranges used in the classifications reflect the following descriptions:

- First Class: the candidate shows excellent skills in reasoning, deductive logic and problem-solving. He/she demonstrates an excellent knowledge of the material, and is able to use that innovatively in unfamiliar contexts.
- Upper Second Class: the candidate shows good or very good skills in reasoning, deductive logic and problem-solving. He/she demonstrates a good or very good knowledge of much of the material.
- Lower Second Class: the candidate shows adequate basic skills in reasoning, deductive logic and problem-solving. He/she demonstrates a sound knowledge of much of the material.
- Third Class: the candidate shows reasonable understanding of at least part of the basic material and some skills in reasoning, deductive logic and problem-solving.
- Pass: the candidate shows some limited grasp of basic material demonstrated by the equivalent of an average of one meaningful attempt at a question on each unit of study. A stronger performance on some papers may compensate for a weaker performance on others.
- Fail: little evidence of competence in the topics examined; the work is likely to show major misunderstanding and confusion, coupled with inaccurate calculations; the answers to questions attempted are likely to be fragmentary only.

## **2 Statistics units and half units**

### **2.1 MS1a: Graphical Models and Inference 16MT**

This unit is not being offered in the 2008-2009 academic year.

### **2.2 MS1b: Statistical Data Mining - 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 Assessment*

This course is assessed by mini-project.

#### *Reading*

C. Bishop, *Neural Networks for Pattern Recognition*, Oxford UP (1995).

D. Hand, H. Mannila, P. Smyth, *Principles of Data Mining*, MIT Press (2001).

I. H. Witten and E. Franke, *Data Mining. Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann (2000).

#### *Further Reading*

B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge UP (1996).

## 2.3 MS2a: Bioinformatics and Computational Biology- 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 an 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 Assessment*

This course is assessed by mini-project.

### *Reading*

C. Semple and M. Steel, *Phylogenetics*, Oxford University Press (2003).

Durbin et al., *Biological Sequence Analysis*, Cambridge University Press (1998).

T. Jiang et al., (editors) *Current Topics in Computational Biology*, MIT Press, (2003).

M. S. Waterman et al., *Computational Genome Analysis: An Introduction*, Springer (2004).

## 2.4 MS2b: Stochastic Models in Mathematical Genetics - 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 Assessment*

1 ½ hour written examination

*Reading*

R. Durrett, *Probability Models for DNA Sequence Evolution*, Springer (2002).

W. J. Ewens, *Mathematical Population Genetics*, 2nd ed, Springer (2004).

J. R. Norris, *Markov Chains*, Cambridge University Press (1999).

M. Slatkin and M. Veuille, *Modern Developments in Theoretical Population Genetics*, Oxford Biology (2002).

S. Tavaré and O. Zeitouni, *Lectures on Probability Theory and Statistics, Ecole d'Eté de Probabilités de Saint-Flour XXXI - 2001*, Lecture Notes in Mathematics 1837.

Springer (2004).

**2.5 MS3b: Lévy Processes and Finance - 16HT**

*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.

*Method of assessment*

1 ½ hour written examination

### *Reading*

J.F.C. Kingman: *Poisson processes*, Oxford University Press (1993), Ch.1-5, 8.

A.E. Kyprianou: *Introductory lectures on fluctuations of Lévy processes with Applications*, Springer (2006), Ch. 1-3, 8-9.

W. Schoutens: *Lévy processes in finance: pricing financial derivatives*, Wiley (2003).

### *Further reading*

J. Bertoin: *Lévy processes*, Cambridge University Press (1996), Sect. 0.1-0.6, I.1, III.1-2, VII.1.

K. Sato: *Lévy processes and infinite divisibility*, Cambridge University Press (1999), Ch. 1-2, 4, 6, 9.

## 2.6 MS4b/C11.1b: **Probabilistic Combinatorics** - 16HT

[In the synopses booklet for Mathematics Part C, this course appears in the Mathematics Department 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 the 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.]

### *Prerequisites*

C11.1a Graph Theory. Part A Probability.

### *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 Assessment*

3 hour written examination also covering C11.1a Graph Theory

*Reading*

N. Alon and J.H. Spencer. *The Probabilistic Method*, Second edition, Wiley, 2000.

*Further reading:*

B. Bollobas, *Random Graphs*, second edition, CUP, 2001.

M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, B. Reed, ed., *Probabilistic Methods for Algorithmic Discrete Mathematics* (Springer, 1998).

S.Janson, T. Luczak and A.Rucinski, *Random Graphs*, John Wiley and Sons, 2000.

M. Mitzenmacher and E. Upfal. *Probability and Computing : Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, New York (NY), 2005.

M. Molloy and B. Reed, *Graph Colouring and the Probabilistic Method* (Springer, 2002).

R. Motwani and P. Raghavan, *Randomized Algorithms* (CUP, 1995).

### 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>

The Mathematics units and half-units that are available are as follows:

- C1.1 (or C1.1a or C1.1b) Gödel's Incompleteness Theorems and Model Theory
- C1.2 (or C1.2a or C1.2b) Analytic Topology and Axiomatic Set Theory
- 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, Banach and C\* Algebras
- C5.1: (or C5.1a or C5.1b or C5.2b) Methods of Functional Analysis
- C6.1a Solid Mechanics
- C6.2b Elasticity and Plasticity
- C6.3 (or C6.3a or C6.3b) Perturbation Methods and Applied Complex Variables
- C6.4a Topics in Fluid Mechanics
- C7.1b Quantum Theory and Quantum Computers
- C7.2b General Relativity I
- 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 2008, 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 2008-2009

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). At least half a unit will be from the schedule of 'Statistics' units for Part C

Please give details of the subjects in which you wish to take classes.  
I wish to take classes in the following subjects: [Please Tick]

- 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,  
1South Parks Road, by the end of week 9 Trinity Term 2008.