UNIVERSITY OF CAMBRIDGE

Part IA Mathematical Biology

Course Handbook 2018 – 2019

For further information on the course, please contact either the Course Organisers, Professor Rufus Johnstone (raj1003@cam.ac.uk) and Dr Nik Cunniffe (njc1001@cam.ac.uk), or the Undergraduate Teaching Administrator in Pathology (teach@path.cam.ac.uk)

Contents

Natural Sciences Tripos Part IA
Mathematical Biology

Mathematical Biology is the first-year Natural Sciences mathematics course designed for biologists. The course teaches a range of skills in mathematical modelling, probability and statistics and simple computer programming. The course is designed to approach these areas from an integrated, biological point of view. A range of mathematical and statistical techniques, including matrix algebra, basic probability and probability distributions, hypothesis testing, regression, ordinary differential equations, local stability analysis, coupled differential equations and modelling are introduced in the context of biological systems.

The course is taught jointly by members of staff from all Departments in the School of Biology (Biochemistry, Genetics, Pathology, Pharmacology, PDN [Physiology, Development and Neuroscience], Plant Sciences, Psychology, Veterinary Medicine and Zoology), as well as staff from the MRC-CBU (the Medical Research Council's Cognition and Brain Sciences Unit).

1. Aims and learning outcomes

Aims

The course has the following aims:

1. to introduce students to the application of mathematical modelling in the analysis of biological systems including populations of molecules, cells and organisms;
2. to show how mathematics, statistics and computing can be used in an integrated way to analyse biological systems;
3. to develop students' skills in algebraic manipulation, the calculus of linear and non-linear differential equations, mathematical modelling, matrix algebra, probability and statistical methods;
4. to introduce students to the use of R for the analysis of biological processes and data, including simple computer programming.

Learning outcomes

At the end of this course, students should:

1. have an enhanced knowledge and understanding of mathematical modelling and statistical methods in the analysis of biological systems;
2. be better able to assess biological inferences that rest on mathematical and statistical arguments;
3. be able to analyse data from experiments and draw sound conclusions about the underlying processes using their understanding of mathematics and statistics;
4. be aware of the use of computers to assist them in studying mathematical functions and carrying out statistical tests.

2. Which students should consider this course?

Mathematical Biology provides a broad base for further studies across all of biology, and should be considered by all biologists, whether their primary interests lie in molecular and cellular

disciplines, physiology, psychology, or in ecology and evolution. The course also provides sufficient mathematical background for certain "physical" subjects in Part IB and Part II, and students reading Mathematical Biology in Part IA have successfully gone on to later studies in, for example, Chemistry and Earth Sciences (although some additional study over the long vacation between first and second year may be required). The course also provides adequate preparation for Part III Systems Biology.

3. Mathematical Biology A and Mathematical Biology B

An A level in Mathematics (or equivalent) is highly recommended for this course. Students with this background – or with an equivalent level of preparation – should take the default version of the course "Mathematical Biology B". Any student who did not study mathematics post-16 will probably study a more restricted version of the course, "Mathematical Biology A". Almost all students who have studied any mathematics post-16 will probably be able to follow Mathematical Biology B, but students with qualifications other than A Level should see below for guidance.

Students taking Mathematical Biology A will attend Block X, a separate series of lectures in the final two weeks of the first (Michaelmas) term, to allow them to follow the material based on calculus in the second (Lent) term. These students will also not be expected to attend the fifth lecture of the (first) lecture series in Block A on probability the first time it is lectured, since this depends on knowledge of calculus. The material will be repeated in a single lecture specifically for Mathematical Biology A students at the end of Michaelmas term.

All students will sit the same written examination in June; **Mathematical Biology A students will have a restricted choice of questions**.

4. Which option is appropriate: Mathematical Biology A or B?

Experience proves that almost all students who have studied mathematics at the post-16 level are adequately prepared to perform well on this course. **In particular, AS Level Mathematics, Scottish Highers, German Abitur, Advanced Placement (AP) Calculus** and a number of other qualifications all provide **sufficient background** to follow the Mathematical Biology B course**,** although in certain cases **a little independent study will be required**. For students entering having done the **International Baccalaureate**, Higher Level Mathematics is highly preferred, but students with **Standard Level Mathematics** will also have **adequate preparation** for the Mathematical Biology B course **(**perhaps again with **some additional self-study** and/or **help from supervisors)**.

Students without one of these backgrounds should have been asked by their College to **self-study** the first four chapters of ***Mathematics for Biological Scientists*** by Aitkin, Broadhurst and Hladky before coming to Cambridge, as well as to have looked at the **additional online resources** at https://www.vle.cam.ac.uk/course/view.php?id=122471. This material – as well as the knowledge of calculus developed by attending the lectures for Block X – is what is required to successfully follow the Mathematical Biology A course.

Colleges will almost certainly wish to arrange for additional supervisions for such students.

Any student concerned about their background – and in particular which of Mathematical Biology A or B would be appropriate – should **discuss this with their Director of Studies** before or soon after arriving in Cambridge. To help students and Director of Studies in this decision, the following table gives a list of topics that will be assumed by the Mathematical Biology B course, as well as the unit of A Level Mathematics in which the topic is covered (all references relate the OCR MEI syllabus 7895, although the particular syllabus used for this comparison is arbitrary, and no A Level syllabus is particularly preferred), and how and when these topics are taught to Mathematical Biology A students.

Students who lack knowledge of most of these topics should almost certainly take the Mathematical Biology A course. However, **if only one or two topics are missing, students can – and probably should –** catch up semi-independently and **still study the Mathematical Biology B course,** perhaps with an additional supervision or two on the missing material. This should be discussed with Directors of Studies. A diagnostic sheet of questions is included at the end of this handbook (Section 17). Students who are able to answer most of these questions should strongly consider Mathematical Biology B, since that choice will ensure the firmest possible foundations for study in Parts IB and II, as well as the widest possible selection of questions in the examination. However, students should talk about any topics that are lacking with their supervisors in the first weeks of the Michaelmas term.

Area of Mathematics	Topic	MEI A Level Module	When is this topic covered in MB-A?
Algebra	Binomial expansion (integer powers)	C1	Online course
	Simultaneous equations with quadratic terms	C1	Online course
	Logarithms	C2	Online course
	Laws of logarithms	C2	Online course
	The exponential function	C3	Online course
	Trigonometric functions: sin, cos, tan	C2	Online course
	Simple trigonometric equations, e.g. sin(x)=1/2	C2	Online course
Calculus	Idea of the gradient of a curve	C2	Block X
	Differentiating: sums and differences of kx^n	C2	Block X
	Finding maxima, minima and inflexion points	C2	Block X
	Product, quotient and chain rules	C3	Block X
	Differentiating: e^{ax+b}, log(ax+b), sin(ax+b), cos(ax+b), (ax+b)^n	C3	Block X
	Integration as reverse of differentiation	C2	Block X
	Integrating: kx^n (when n is not -1)	C2	Block X
	Integration as area under a graph	C2	Block X
	Integrating: e^{ax+b}, 1/(ax+b), sin(ax+b), cos(ax+b), (ax+b)^n	C3	Block X
Curve sketching	Sketching e^x, log(x), sin(x), cos(x), tan(x)	C2 & C3	Online course
	Transformations of graphs: given a curve f(x), sketching f(x) + a, f(x+a), f(ax), af(x)	C1 & C2	Online course

5. Components of the course

Lectures. There are three lectures per week, held at 9am on Tuesday, Thursday and Saturday. All lectures for Mathematical Biology B students are held in the Main Lecture Theatre of the Department of Zoology on the New Museums Site. Mathematical Biology A students are expected to attend almost all of these lectures, but should note that their separate lectures for Block X will be in the Mill Lane Lecture Rooms on Mill Lane (details of the precise room will be circulated at the time).

Practicals. It is increasingly important for biologists to have skills in using computers. We therefore teach the statistical programming language, R. The goal is that by the end of the course, all students should be able to work comfortably with data, and be able to do some simple computer programming. The practical classes are held in the Titan Teaching Rooms, and you will do a single 1.25 hour practical every week on a Thursday afternoon (at one of 2pm, 3.30pm or 4.45pm, depending on your timetable). If for some reason you cannot attend your assigned practical time please email teach@path.cam.ac.uk with details. It may be possible to accommodate you in another session.

Supervisions. Your College Director of Studies will organise supervisions on your behalf. Supervisions are one-hour sessions with a lecturer or post-graduate/post-doctorate research worker. This offers the opportunity to discuss course work and iron out problems. For this course, supervisions almost always will consist of going through the solutions to the questions on Examples Sheets that are handed out in the Lectures, although supervisors should also be able to clear up any problems students encounter in lectures or other study.

6. Assessment and examinations

The course is assessed by one 3-hour written examination in the Easter Term (worth 80% of the final mark), and by assessment of coursework exercises based on the practical component of the course (worth 20% of the final mark).

For the written examination, Section A will consist of questions on material covered in the first half of the Michaelmas Term (Block A); Section B of questions on material covered in the second half of the Michaelmas Term (Block B); Section C of questions on material covered in the first half of the Lent Term (Block C); Section D of questions on material covered in the second half of the Lent Term (Block D); and Section E of questions on material covered in the Easter Term (Block E). **There will be two questions in each of Sections A to D, with three questions in Section E. Candidates will be required to answer eight questions, selecting at least one question from each section, and no more than two questions from any section (i.e. all three questions in Section E cannot be attempted).**

At most one of the questions in Section B will require material covered in the last six lectures of the Michaelmas term (i.e. at most one of the questions in Section B will be inaccessible to students who do Mathematical Biology A; see below). This question will be clearly marked on the examination paper. **Note that (for students doing Mathematical Biology A) Block X is not**

explicitly assessed on the examination (although students' knowledge of calculus is implicitly assessed via the questions on Blocks C, D and E).

The form of the examination has therefore been slightly altered from 2017-18 (last year there were only two questions in Section E, but the paper otherwise was of an identical format).

For the practical assessment, candidates will be required to submit two written exercises based on material covered in the practical classes. One written exercise will be set on the material covered during the Michaelmas term and one written exercise will be set on the material covered during the Lent term. The corresponding exercises must be submitted by the Thursday in week 3 of the following term: this year, the hand-in dates are therefore Thursday 31st January 2019 and Thursday 9th May 2019. Details of how to submit the work – which will be done via a Moodle site – will be provided at the time the work is set.

There are three types of examiner for the course: a Senior Examiner, an Assistant Senior Examiner (who was Senior Examiner in the preceding year) and a Junior Examiner (who will be Senior Examiner next year). The Senior Examiner for 2018-19 is Dr Bill Broadhurst, Biochemistry. There are two Assistant Senior Examiners, Dr Nik Cunniffe, Plant Sciences and Prof. Rufus Johnstone, Zoology. The Junior Examiner is Professor Ewa Paluch, PDN. The Senior Examiner(s) are responsible for the administrative side of running the examination, although individual Lecturers generally set examination questions in consultation with the Senior Examiner. The Examiners also collectively set coursework exercises based on the practical component of the course.

7. Online resources

Online resources are provided through Moodle. In particular, lecture handouts and the slides shown during lectures will be made available via Moodle. Examples sheets will also be put online via Moodle, with worked solutions to these examples sheets provided after some delay (i.e. once supervisions on the material have been completed). The handouts used in practical classes, as well as R markdown (you will learn what this is in the practical classes) giving solutions to any exercises, will also be posted on Moodle.

You will be subscribed to the Moodle site automatically as part of the NST IA subject choice procedures. Your Hermes user name and password will allow you to access Moodle from outside Cambridge during the vacations.

8. Student feedback

Mathematical Biology is run jointly by a number of Departments. Responsibility for what is taught on the course rests with the Management Committee, which consists of all lecturers involved in teaching the course, as well as a representative elected from among the supervisors. The same set of people – augmented by the four undergraduate student representatives (see below) – also sit on the Consultative Committee. This Committee discusses feedback on the content of the lectures and practical classes, seeking feedback in two ways:

- replies to questionnaires;

- collation of student views by the student representatives.

Questionnaires. As the course progresses, you will be asked to fill in electronic questionnaires about the course. It is vital to fill in these questionnaires, since the answers help us understand your experience of the course, and make any changes that might be required.

Consultative Committee. You will be asked to elect your own representatives during the Michaelmas Term. It is important that you participate in the nomination and election of your representatives and use them to make your views known to the Consultative Committee, who meet towards the end of each term.

You may of course direct comments directly to the person who is currently lecturing the course, as well as the Course Organiser, Undergraduate Teaching Administrator or to any other member of teaching staff whenever you wish.

9. Formula book

A formula book will be provided for the examination, and will be distributed for reference in the first lecture at the start of the course. However, please note that this book only contains formulae – and statistical tables – relevant to the material covered in Block B. **Students will be expected to remember any pertinent formulae and/or mathematical methods from the lectures for Block A, C, D and E** (supervisors may wish to note this is different to the version of the course up to and including the 2016-17 academic year).

10. Downloading R and R Studio

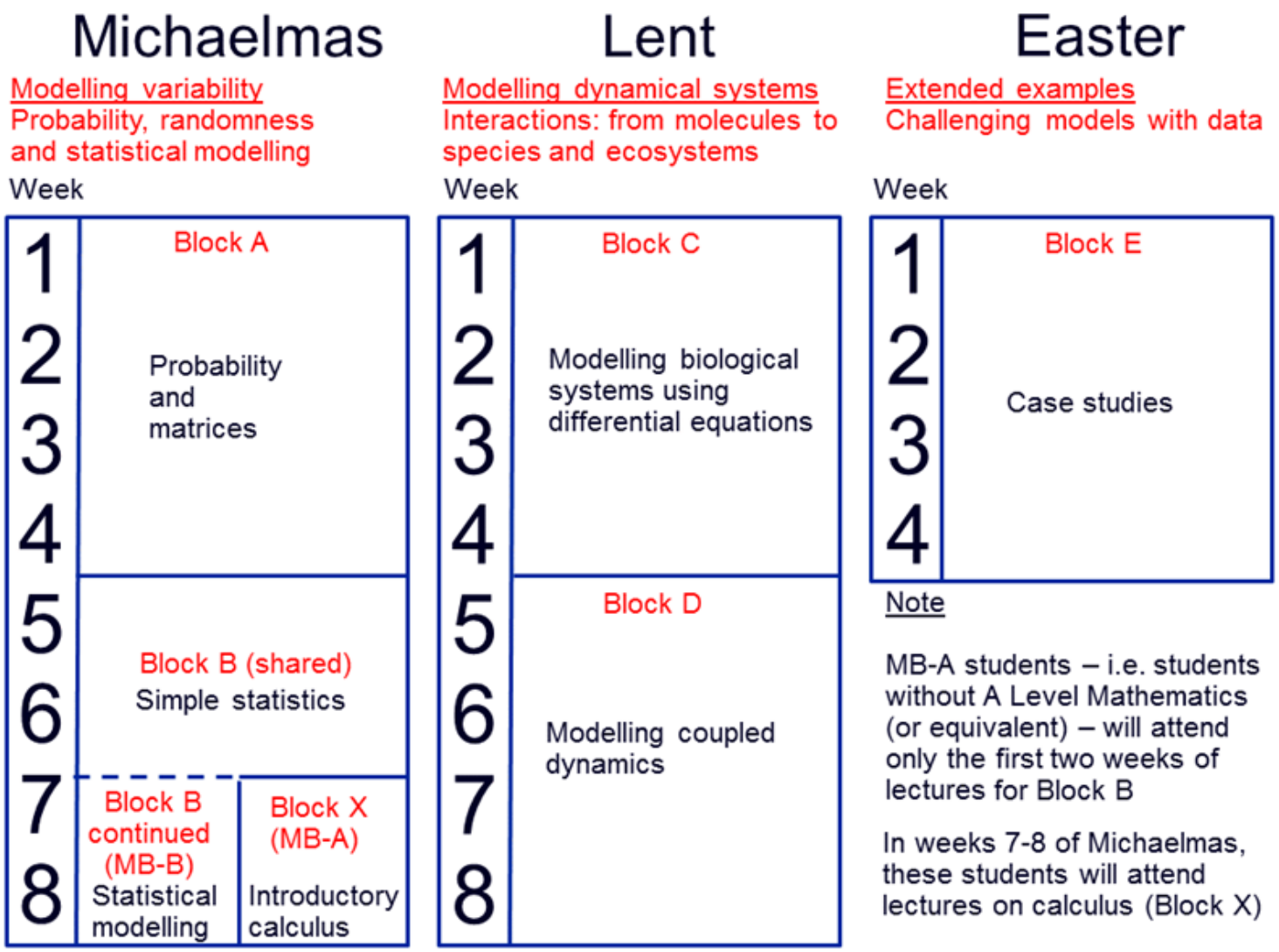
Practical classes teach the statistical programming language, R. **The goal is that by the end of the course, all students should be able to work comfortably with data, and be able to do some simple computer programming.** Although the practical classes are self-contained, in the past students have found it very helpful to install a copy of the software on their own computers. Having a copy of the software installed on your own computer might also be helpful in doing the assessed exercises.

R can be installed – for a number of platforms, including Windows, Mac, Unix and Linux – by downloading from https://www.r-project.org/ (where there are full instructions). We recommend and teach the use of the integrated development environment R Studio: this is available for download from https://www.rstudio.com/products/rstudio/

If you do not have a computer, note that R and R Studio are installed on all Managed Cluster Service (MCS) computers installed by the University Information Service: this will include computers in College libraries and computer rooms.

11. Overall structure of the course

A schematic showing the overall structure of the course is given on the following page. More detail on the contents of the individual blocks are given later in this handbook.


Michaelmas
Modelling variability
Probability, randomness and statistical modelling
Week
1
2
3
4
5
6
7
8
Block A
Probability and matrices
Block B (shared)
Simple statistics
Block B continued (MB-B)
Statistical modelling
Block X (MB-A)
Introductory calculus
Lent
Modelling dynamical systems
Interactions: from molecules to species and ecosystems
Week
1
2
3
4
5
6
7
8
Block C
Modelling biological systems using differential equations
Block D
Modelling coupled dynamics
Easter
Extended examples
Challenging models with data
Week
1
2
3
4
Block E
Case studies
Note
MB-A students – i.e. students without A Level Mathematics (or equivalent) – will attend only the first two weeks of lectures for Block B
In weeks 7-8 of Michaelmas, these students will attend lectures on calculus (Block X)

12. Lecture timetable

MICHAELMAS TERM			
1-6	4,6,9,11,13,16 Oct	Introduction to probability	Dr Ioanna Mela (teaching for Pharmacology)
7-12	18,20,23,25,27, 30 Oct	Matrix algebra	Dr Aylwyn Scally (Genetics)
13-15	1,3,6 Nov	Introduction to statistics	Dr Kristian Franze (PDN)
16-18	8,10,13 Nov	Statistics: Linear models I	Prof. Andrea Manica (Zoology)
MB-A			
19-24	15,17,20,22,24,27 Nov	Introduction to basic calculus	Dr Bill Broadhurst (Biochemistry)
-	10am, 27 Nov	Repeat of probability lecture 5 for MB-A students	Dr Ioanna Mela (teaching for Pharmacology)
MB-B			
19-21	15,17,20 Nov	Statistics: Linear models II	Prof. Andrea Manica (Zoology)
21-24	22,24,27 Nov	Generalised linear models	Dr Stephen Sawiak (teaching for Psychology)
LENT TERM			
25-36	17,19,22,24,26,29, 31 Jan and 2,5,7,9,12 Feb	Modelling biological systems using differential equations	Dr Nik Cunniffe (Plant Sciences)
37-42	14,16,19,21,23,26 Feb	Coupled non-linear equations	Dr Marta Correia (MRC-CBU, teaching for SBS)
43-48	28 Feb and 2,5,7,9,12 Mar	Models of interacting species	Dr Olivier Restif (Veterinary Medicine)
EASTER TERM			
49-52	25,27,30 Apr and 2 May	Introduction to bioinformatics	Dr Andrew Firth (Pathology)
53-56	4,7,9,11 May	Modelling reaction kinetics	Dr Bill Broadhurst (Biochemistry)
57-60	14,16,18,21 May	Evolutionary modelling	Prof. Rufus Johnstone (Zoology)

13. Detailed content of individual blocks of lectures

MICHAELMAS TERM. BLOCK A. PROBABILITY AND MATRICES

Dr Ioanna Mela
(Pharmacology)
im337@cam.ac.uk

Dr Aylwyn Scally
(Genetics)
aos21@cam.ac.uk

Synopsis

Part 1 introduces concepts in probability, starting with how probabilities of single and multiple events are defined and ideas of independence and dependence. Key to many probability calculations is the mathematics of counting, or combinatorics, which enables us to enumerate the number of ways in which an event may occur. The lectures also introduce probability distributions, including certain special distributions used to represent random processes in the real world. These enable us to model biological phenomena and make inferences about the factors involved in an experiment. Part 2 concerns mathematical objects called matrices and how to carry out calculations involving them. Matrices are used widely in statistics and biology and are the basis for many data analyses. We introduce the algebra of matrices and study their properties as far as eigenvalues and eigenvectors, which enable us to evaluate the result when a matrix is repeatedly applied to a vector.

Details of individual lectures

Block A, Part 1 (Lectures 1-6). Introduction to probability (Ioanna Mela)

1. Sample spaces and events
2. Combinations and permutations
3. Discrete probability distributions
4. Continuous probability distributions
5. Expectation (Lecture is repeated for MB-A students at end of the term)
6. Worked examples

Block A, Part 2 (Lectures 7-12). Matrix algebra (Aylwyn Scally)

7. Basic matrix algebra
8. Linear systems, including Gaussian elimination, matrix inversion and determinants.
9. Eigenvalues and eigenvectors (Part One)
10. Eigenvalues and eigenvectors (Part Two)
11. Powers of matrices

MICHAELMAS TERM. BLOCK B. STATISTICS AND STATISTICAL MODELLING

(NB. MB-A students attend only the first six lectures in Block B, then switch to Block X)

Dr Kristian Franze
(PDN)
kf284@cam.ac.uk

Prof. Andrea Manica
(Zoology)
am315@cam.ac.uk

Dr Stephen Sawiak
(Psychology)
sjs80@cam.ac.uk

Synopsis

These lectures cover a broad range of statistical techniques used by biologists to analyse data. We will discuss how to pose biological questions as statistical hypotheses that can be formally tested. We will start with simpler tests, such as t-tests and χ^2, and progress to explore linear models that allow us to account for multiple predictors simultaneously. Besides understanding the logic and mechanics of tests, we will focus on how to test whether the assumptions of each test are met, and how to proceed in the – surprisingly common – situations when they are not. The lectures will use output from the statistical language R, providing the theoretical background to correctly interpret the outcome of different analyses when used in real life.

Details of individual lectures

Block B, Part 1 (Lectures 13-15). Introduction to statistics (Kristian Franze)

12. Introduction to statistics: hypothesis testing, binomial tests, Chi squared
13. t-tests & confidence intervals
14. Two sample t-tests and non-parametric tests

Block B, Part 2 (Lectures 16-18). Statistics: Linear models I (Andrea Manica)

15. One-way analysis of variance (ANOVA)
16. Assumptions of ANOVA, diagnostic plots, Tukey HSD's & Kruskal-Wallis test
17. Regression and its assumptions

Block B, Part 3 (Lectures 19-21). Statistics: Linear models II (Andrea Manica)

18. Linear model notation and interactions
19. Interactions and model simplification via backwards stepwise elimination
20. Model selection via the Akaike Interaction Criterion

Block B, Part 4 (Lectures 22-24). Generalised linear models (Stephen Sawiak)

21. Binary data: limitations of linear models, link functions and binomial errors
22. Count data: models with a Poisson error structure
23. Dealing with challenging data

MICHAELMAS TERM. BLOCK X. INTRODUCTION TO BASIC CALCULUS

(NB. These lectures are to be attended by MB-A students only)

Dr Bill Broadhurst
(Biochemistry)
rwb1002@cam.ac.uk

Synopsis

These lectures introduce the basic principles of calculus, which is helpful for predicting how biological systems can change over time or over other variables, such as distance. The purpose of Block X is solely to provide those who have not studied mathematics at the post-16 level with the concepts they need to handle the material covered in Block C in Lent term. We therefore expect the majority of students to attend the lectures in Parts 3 and 4 of Block B instead.

Block X, starts with an introduction to differentiation, providing a useful way to think about rates of change and why curves are curved. This area has several useful applications, including how to sketch the shape of an unknown function, how to make approximations and how to handle experimental errors. Next comes an introduction to integration, an operation that in effect undoes the consequences of differentiation and enables the area under a curve to be calculated. The final section explains what differential equations are, how they can be used to model biological phenomena such as changes in population, how simple examples can be solved, and the importance of boundary conditions for obtaining a unique solution.

Details of individual lectures

Block X (Lectures 19-24). Introduction to basic calculus (Bill Broadhurst)

19. Curves, gradients and differentiation from first principles
20. Maxima, minima and curve sketching
21. How to differentiate more complex functions
22. Integration and the area under a curve
23. How to integrate more complex functions
24. Introduction to differential equations

LENT TERM. BLOCK C. MODELLING USING DIFFERENTIAL EQUATIONS

Dr Nik Cunniffe
(Plant Sciences)
njc1001@cam.ac.uk

Synopsis

These lectures show how a range of biological phenomena can be modelled using differential equations, focusing on cases in which the value of a single state variable is changing continuously in time. Examples are largely drawn from population dynamics, although the techniques are relevant much more broadly (see Blocks D and E). The main mathematical skill is solving differential equations via separation of variables. However, qualitative analyses – finding model equilibria and examining long-term stability, as well as using the direction field to sketch solution curves – can often be equally revealing. Such methods can be particularly useful since they allow more complex systems to be investigated without any significant additional conceptual difficulty. These methods also generalise more readily to situations in which the value of multiple state variables are changing in time in a coupled way, for example chemical reactions involving multiple chemical species (see Block E, Part2), or ecological interactions involving more than one biological species with population sizes that depend on each other (see Block D).

Details of individual lectures

Block C (Lectures 25-36). Modelling using differential equations (Nik Cunniffe)

25. Techniques I. Calculus, exponentials, sketching and differential equations
26. The exponential model
27. The monomolecular model
28. Techniques II. Methods for solving differential equations
29. The logistic model I. Setting up and solving the model
30. The logistic model II. Interpreting the solution of the model
31. Techniques III. Equilibria, stability and qualitative analysis
32. The von Bertalanffy model
33. Rate parameters that depend on time
34. Techniques IV. Series approximations of functions
35. An analytic test for stability of an equilibrium
36. Extending and linking the basic models

LENT TERM. BLOCK D. MODELLING COUPLED DYNAMICS

Dr Marta Correia
(MRC CBU)
mmc43@cam.ac.uk

Dr Olivier Restif
(Veterinary Medicine)
or226@cam.ac.uk

Synopsis

These lectures extend the mathematical methods covered in Block C to allow analysis of more complicated biological models (in particular the dynamics of populations that are coupled together). The main focus is an examination of the dynamics of linear and non-linear coupled differential equations, both analytically and by graphical "phase plane" techniques. In the second half of Block D we use these methods to look at various coupled population systems, including predator-prey systems, competition within and between species, and the mathematical modelling of epidemics in plants, humans and animals.

Details of individual lectures

Block D, Part 1 (Lectures 37-42). Coupled non-linear equations (Marta Correia)

37. Introduction to systems of linear first order differential equations
38. Analytical solution to systems of linear first order differential equations
39. Null-clines, equilibria and stability
40. Functions of more than one variable: partial differentiation and Taylor approximation
41. Solving systems of non-linear first order differential equations
42. Systems of non-linear first order differential equations: examples and applications

Block D, Part 2 (Lectures 43-48). Modelling interacting species (Olivier Restif)

43. Growth models revisited
44. Interacting species: predation models
45. Interacting species: competition models
46. Viral infection dynamics
47. Simple epidemic models
48. Advanced epidemic models

EASTER TERM. BLOCK E. CASE STUDIES

Dr Bill Broadhurst
(Biochemistry)
rwb1002@cam.ac.uk

Dr Andrew Firth
(Pathology)
aef24@cam.ac.uk

Prof. Rufus Johnstone
(Zoology)
raj1003@cam.ac.uk

Synopsis

Part 1 introduces methods for the analysis of biological sequence data. The lectures will discuss existing software and algorithms with an overview of underlying theory. Part 2 makes use of some of the mathematical tools discussed in Lent term to gain insight into a range of widely applicable chemical and biochemical processes, from radioactive decay, to reactions catalysed by enzymes and metabolic reaction networks. Part 3 looks at models of animal behaviour based on the idea of selection as a fitness-maximising process. Beginning with straightforward applications of optimality, we move on to the concept of evolutionary stability and simple game theoretical models of social behaviour, focusing in particular on the evolution of parental care as an illustrative case.

Details of individual lectures

Block E, Part 1 (Lectures 49-52). Introduction to bioinformatics (Andrew Firth)

49. Introduction: Dynamic Programming
50. Pairwise and multiple sequence alignment
51. Homology search: Sequence evolution
52. Phylogenetic analysis

Block E, Part 2 (Lectures 53-56). Modelling reaction kinetics (Bill Broadhurst)

53. Unimolecular, bimolecular and reversible reactions
54. Sequential reactions and the steady state approximation
55. Enzyme kinetics and inhibition
56. Extended examples

Block E, Part 3 (Lectures 57-60). Evolutionary modelling (Rufus Johnstone)

57. Optimality
58. Game theory
59. Cyclical dynamics
60. Extended examples

14. Practical timetable

All practical classes are in the Titan Teaching Rooms on the New Museums site. Practical sessions are repeated at three times per week: 2pm, 3.30pm and 4.45pm. Please ensure you bring your MCS user-id and password to all sessions.

Michaelmas term

1	4th Oct 2018	Introduction to R & using variables	Dr Ioanna Mela (Pharmacology)
2	11th Oct 2018	Vectors, built-in functions & R markdown	Dr Ioanna Mela (Pharmacology)
3	18th Oct 2018	Scripts, random numbers & simple plots	Dr Alywyn Scally (Genetics)
4	25th Oct 2018	Matrices & the apply function	Dr Alywyn Scally (Genetics)
5	1st Nov 2018	File handling & data frames	Dr Kristian Franze (PDN)
6	8th Nov 2018	Basic statistics & simple plotting	Prof Andrea Manica (Zoology)
7	15th Nov 2018	ANOVA & advanced plots	Prof Andrea Manica (Zoology)
8	22nd Nov 2018	R Markdown & linear statistical models	Dr Steve Sawiak (Psychology)

Lent term

-	17th Jan 2019	Optional drop-in session to obtain help with assessed exercises	All lecturers
9	24th Jan 2019	For loops	Dr Nik Cunniffe (Plant Sciences)
10	31st Jan 2019	Nested loops, basic logical operators & while loops	Dr Nik Cunniffe (Plant Sciences)
11	7th Feb 2019	Further logical operators & conditionals	Dr Nik Cunniffe (Plant Sciences)
12	14th Feb 2019	User-defined functions	Dr Marta Correia (SBS)
13	21st Feb 2019	Programming, debugging & error handling	Dr Marta Correia (SBS)
14	28th Feb 2019	Euler's algorithm for differential equations	Dr Olivier Restif (Veterinary Medicine)
15	7th Mar 2019	Solving differential equations	Dr Olivier Restif (Veterinary Medicine)

Easter term

Note there are only three practical classes in the Easter term.

-	25th Apr 2019	Optional drop-in session to obtain help with assessed exercises	All lecturers

16	2nd May 2019	Bioinformatics	Dr Andrew Firth (Pathology)
17	9th May 2019	Reaction kinetics	Dr Bill Broadhurst (Biochemistry)

15. Relationship to previous version of the course

The Mathematical Biology course was totally redesigned for the 2017-2018 academic year. However, approximately 75% of the material was covered in the old course. The following is largely intended for supervisors, and shows the mapping from the old to the new course

- Block A
  - Part 1. Previously covered by Aylwyn Scally's lectures in Michaelmas
    - Very similar content to the old course
  - Part 2. Previously covered by Nik Cunniffe's lectures in Easter.
    - Final lecture on Leslie matrices has been dropped. Material on Gaussian elimination and matrix inversion has been added.
- Block B
  - Part 1. Previously covered by Rufus Johnstone's lectures in Lent
    - Different emphasis to the old course (fewer hand calculations)
  - Parts 2 and 3. Previously covered by Andrea Manica's lectures in Lent
    - Very similar content to old course, but final lecture moved into Part 4
  - Part 4.
    - Approximately one lecture covered by Andrea, but rest entirely new
- Block C
  - Previously covered by Nik Cunniffe's lectures in Michaelmas
    - Similar content, but dropped delay parameters, dimensional analysis, discrete time models and integrating factors; added Taylor series, analytic stability tests, explicit coverage of extending and linking the models, and explicit coverage of Holling responses for predation
- Block D
  - Part 1. Previously covered by John Trapp's lectures in Lent
    - Dropped second order differential equations and functions of two variables; focus on matrix version of linear first order systems
  - Part 2. Previously (largely) covered by Colin Russell's lectures in Lent
    - Very similar content, but added some new material
- Block E
  - All new, although Rufus Johnstone's lectures on Evolutionary Modelling were covered in the old version of Mathematical Biology until about 2011 or so
- Block X
  - All material was previously covered by Bill Broadhurst's lectures in Michaelmas in the old course Elementary Mathematics for Biologists

16. Relevance of old examination questions

The 2017-18 examination paper is very similar in format to the current course, with the only exception being that there were questions on only two of the three topics covered in Block E.

It might be helpful to note in revision the following table, which shows which old Tripos examination questions from years before 2017-18 test material that is still covered in the course. Capital letters in the table show the block to which the question is relevant. Any lower case letters denote parts of the question which are no longer accessible, since the material in the part(s) with lower case letters have been removed from the syllabus. For example Question A1 on the 2013 paper is marked C/be; this means that the question is relevant to Block C in the new course, but material in parts b and e should be ignored (in this case since these parts focus on dimensional analysis, which has been deleted from the syllabus). In studying the old examination questions, note that the "A questions" on old papers were intended to take up to 15 minutes, and the "B questions" up to 30 minutes: all questions on the new examination papers will be of length intermediate between these. You should also note that – as outlined on the previous page – some parts of the new course are not represented on the old exam papers.

	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	B6
2017	C	A	-	D	B	-	C/ae	A	D	B	D	A
2016	C	A	-	D	B	A	C	-	D	B	D	-
2015	-	A/ef	-	B	-	D	C	A	D	B	A	D
2014	C	A	-	B	-	D	C/e	A	D	B	A	D
2013	C/be	-	-	D	A	D	C	-	D	B	-	D
2012	C	-	D/b	-	A	D	C	-	D	B	-	D
2011	C/a	-	-	B	A	D	C	-	D	B	A	D
2010	C	-	D	B	-	D	C	-	D	B	-	D

17. Diagnostic questions to inform choice of Mathematical Biology A or B

(Note. Students who are not completely confident that they have the requisite background should look over these questions and discuss with their supervisors about any gaps as early as possible in the Michaelmas term. One or two gaps can almost certainly be quickly filled in while doing the Mathematical Biology B course. However, if more topics have not been covered, it might be that Mathematical Biology A is the more appropriate choice).

Questions

Qu 1. What is $(a+b)^6$? (NB. you are not expected to multiply six sets of brackets!)

Qu 2. Solve the following simultaneous equations $y = x^2 - 9x + 20$ and $x = 1+y$.

Qu 3. Solve the following equations for x: i) $5^x=4$; ii) $\log_2(x+1)-\log_2(x^2-1) = 2$; iii) $\sin(x) = 0.5$.

Qu 4. Sketch the following graphs: i) $y=2e^{2x}$; ii) $y=\sin(3x)$; iii) $y = 7\sin(3x+2)$

Qu 5. Differentiate with respect to x: i) $y=10x^2+4$; ii) $y=x^3-3x^{10}$

Qu 6. Find the turning points of $y = 3x^4 + 4x^3 - 12x^2$, and give a rough sketch of the curve.

Qu 7. Differentiate with respect to x: i) $y=e^x\sin(x)$; ii) $x/(1+x)$; iii) $\log(x^3+7x)$; iv) $(\sin(2x) + x^3)^{10}$.

Qu 8. Integrate with respect to x: i) x^{10}; ii) $\sin(3x+8)$; iii) $(1+2x)^5$; iv) $1/(x+1)$

Answers

Qu 1. $(a+b)^6 = a^6 + 6a^5b + 15a^4b^2 + 20a^3b^3 + 15a^2b^4 + 6ab^5 + b^6$.

Qu 2. $x = 3$ and $y = 2$ or $x = 7$ and $y = 6$.

Qu 3. i) x = log(4)/log(5) [base doesn't matter]; ii) x = 5/4; iii) $x = \frac{\pi}{6} + 2\pi n, \frac{5\pi}{6} + 2\pi n$.

Qu 4. Sketches omitted, but can be verified (e.g.) using Wolfram Alpha.

Qu 5. i) $dy/dx = 20x$; ii) $dy/dx = 3x^2 - 30x^9$.

Qu 6. (1,-5) (local minimum); (0,0) (local maximum) and (-2,-32) (local minimum).

Qu 7. i) $e^x(\sin(x)+\cos(x))$; ii) $1/(1+x)^2$; iii) $(3x^2+7)/(x^3+7x)$; iv) $10(2\cos(2x)+3x^2)(\sin(2x)+x^3)^9$.

Qu 8. i) $x^{11}/11 + C$; ii) $-\cos(3x+8)/3 + C$; iii) $(1+2x)^6/12 + C$; iv) $\log(x+1) + C$.

18. Mock-up of a written examination paper

NATURAL SCIENCES TRIPOS	**Part IA**
June 2019 (Sample Exam)	9 to 12

MATHEMATICAL BIOLOGY

*You must answer **eight** questions.*

*You must answer at least **one question** from **each of Sections A to E**.*

*You must **not** answer all three questions from **Section E***

*You must begin each answer on a **separate** sheet.*

*Attach a **separate** cover sheet to **each** question.*

The question in Section B that is marked with an asterisk (*) requires knowledge of the last six lectures in the Michaelmas term.

Indicative proportions of marks for each part of the questions are given.

It does not matter whether you write on only one side of the paper or on both sides.

STATIONERY REQUIREMENTS
Script Paper
Rough Work Pads
Blue Coversheets
Tags

SPECIAL REQUIREMENTS
Formulae Booklet
Approved Calculators Allowed

You may not start to read the questions printed on the subsequent pages of this question paper until you have been instructed that you may do so by the Invigilator

SECTION A

A1

(a) A bird of prey hunts mice. The chance of success for each attempt by the bird to catch a mouse is 0.6. The bird makes four attempts. Find the probability of each of the following outcomes.

i. Three successful attempts.
ii. No successful attempts.
iii. At most two successful attempts. *[~30% marks]*

(b) A second bird of prey hunts with a probability of success p=0.8. Two graduate students are studying the behaviour of this particular bird and devise the following game: Student A pays 50p to Student B every time the bird makes a successful attempt and gets 80p from Student B every time the bird fails to catch a mouse. They play the game over twenty attempted hunts by the bird. What is the expected value of the earnings of Student B?

[~30% marks]

(c) The same bird of prey also hunts squirrels. In its natural habitat, it can find both red and grey squirrels. The grey squirrels make up 80% of the squirrel population, while red squirrels make up 20%. The dynamics of this squirrel population are being studied and for this purpose the animals are being tagged. 92% of all the red squirrels have been tagged and 63% of the grey. A tagged squirrel is caught by the bird of prey. What is the chance that the squirrel was red?

[~40% marks]

A2

Consider the following matrix

$$A = \begin{pmatrix} 4 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 2 \end{pmatrix}$$

(a) Find $A\mathbf{v}^T$ where $\mathbf{v}$ = (1,3,-2). *[~5% marks]*

(b) Find det(A). *[~15% marks]*

(c) Find the characteristic equation of A, and hence find its eigenvalues. *[~40% marks]*

(d) Find a set of eigenvectors for A, and hence sketch its invariant lines when viewed as a linear transformation. *[~40% marks]*

SECTION B

B3 A biologist suspects that a ladybird population may be subject to infection by a male-killing bacterium, which gives rise to a female-biased sex ratio. To test this idea, he samples 10 individuals from the population and records their sex. He plans to use a one-tailed binomial test (with a significance level of α = 0.05) to evaluate whether the sex ratio is significantly skewed towards females.

(a) What is the minimum number of females (out of the 10 individuals sampled) that would cause the biologist to reject the null hypothesis of a 1:1 population sex ratio? *[~20% marks]*

(b) Suppose that the sex ratio is in fact 1:1. What is the probability of obtaining sufficient females in the sample to mistakenly reject the null hypothesis (a Type I error)? *[~10% marks]*

(c) Suppose that the sex ratio is in fact 3:1 (females:males). Calculate the probability of the biologist obtaining so few females that he mistakenly fails to reject the null hypothesis (a Type II error)? *[~20% marks]*

(d) How might the risk of both types of error be reduced? *[~10% marks]*

A statistics lecturer compares the heights (in cm) of 15 male students from three colleges: University Hall, Valance Mary Hall and Katharine Hall.

University Hall	Valance Mary Hall	Katharine Hall
177	172	172
178	179	180
180	178	182
184	180	176
182	185	187

$$\bar{y}_{University} = 180.2\,\text{cm};\ \bar{y}_{Valance\,Mary} = 178.8\,\text{cm};\ \bar{y}_{Katharine} = 179.4\,\text{cm}$$

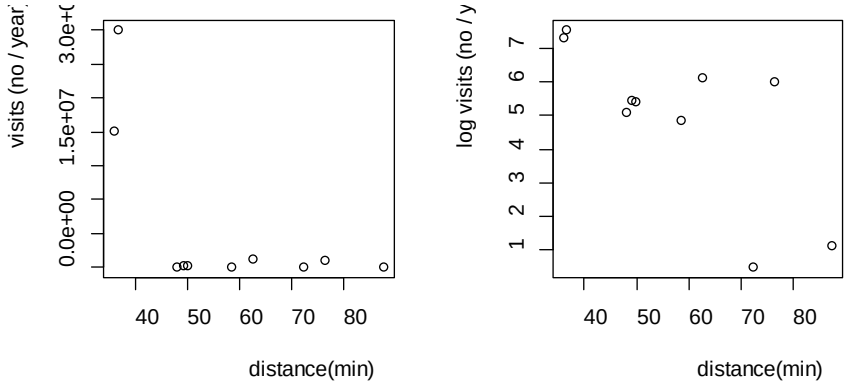
$$\sum y_i = 2692\,\text{cm};\qquad \sum (y_i - \bar{y})^2 = 255.73\,\text{cm}^2.$$

(e) Is there a significant difference between heights of male students at these Colleges? (Please show working and state null and alternative hypotheses) *[~30% marks]*

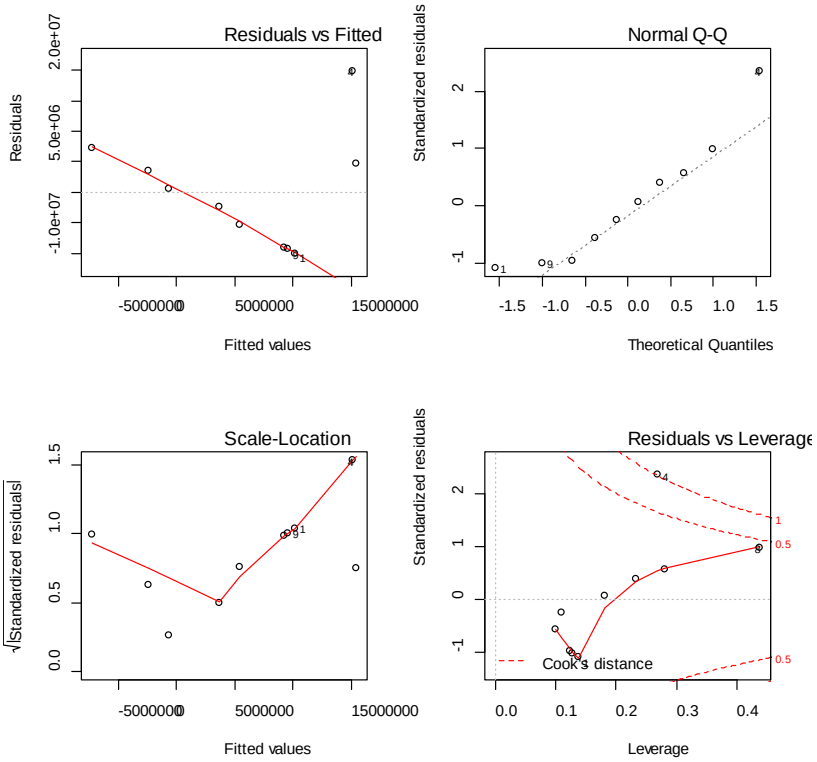
(f) Provide one or two sentences that could be used to report your analysis in a scientific paper. *[~10% marks]*

B4* (Requires knowledge of the last six lectures of the Michaelmas term)

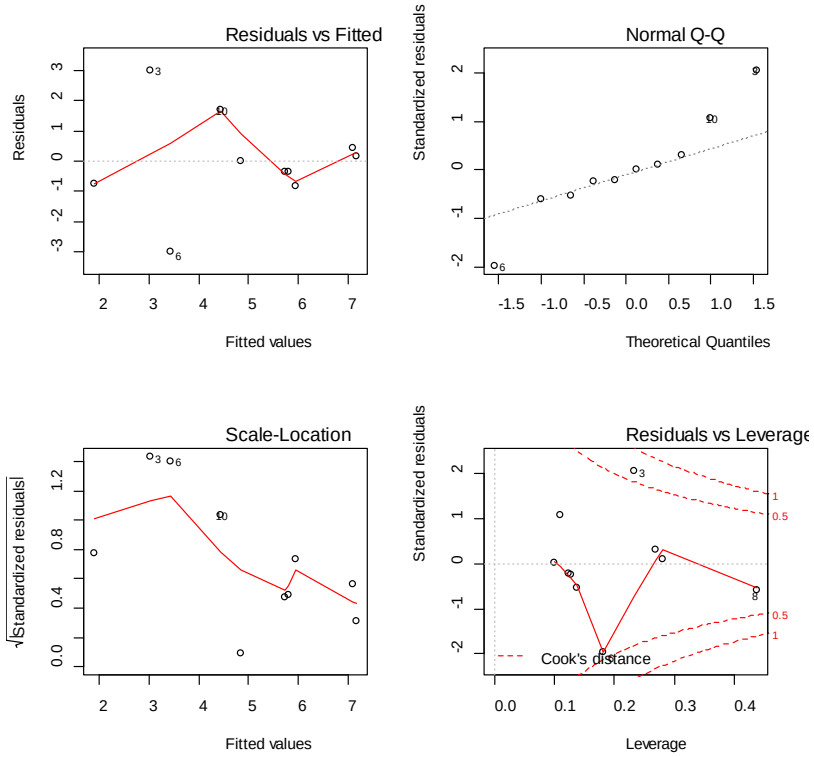
A conservation biologist is interested in testing whether accessibility affects the number of visitors to nature reserves. She collects visitor data (number of visits/year) for 10 reserves, for which she also measures driving distance (minutes) to the nearest town. She plots both the raw data as well as log transforming visit numbers:



After inspecting the diagnostic plots for a regression analysis on the raw data,



she decides to carry out her regression analysis on the log transformed data, and obtains the diagnostic plots shown in the next page.



(a) Was she justified in deciding to use the transformed data? If yes, did the transformation improve the validity of the model's assumptions?

[~15% marks]

When she carries out her regression on the log transformed data, she obtains the following ANOVA table.

	SS	df	MS	F
Model	27.904	**??**	**??**	**??**
Error	**??**	**??**	**??**	
Total	49.712	9		

(b) Complete the table (missing items are denoted by **??**), and test whether distance from the nearest town is a predictor of the number of visits.

[~20% marks]

(c) Write a sentence to summarise your results in a scientific paper.

[~15% marks]

An analysis of whether weight (in kg) and sex of patients affects their likelihood of contracting an infection while on hospital returns the following output:

```
Call:
glm(formula = infected ~ sex + weight + sex:weight,
      family = binomial(logit), data = Dataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6236  -0.5969  -0.3051  -0.2450   2.3101

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)         1.18270    0.92802   1.274   0.2025
sex[T.male]        -0.86963    1.29470  -0.672   0.5018
weight             -0.17608    0.09220  -1.910   0.0562 .
sex[T.male]:weight -0.04771    0.12700  -0.376   0.7072
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 83.234  on 80  degrees of freedom
Residual deviance: 63.645  on 77  degrees of freedom
AIC: 71.645
```

(d) What is the likelihood of contracting an infection for a male patient weighting 47 kg?

[~10% marks]

Dropping the interaction we get:

```
Analysis of Deviance Table

Model 1: infected ~ weight + sex
Model 2: infected ~ weight * sex
Resid. Df Resid. Dev Df Deviance
1        78     63.785
2        77     63.645  1    0.141
```

Dropping the main factors from the additive model, we get:

```
Analysis of Deviance Table

Model 1: infected ~ sex
Model 2: infected ~ weight + sex
Resid. Df Resid. Dev Df Deviance
1        79     74.956
2        78     63.785  1   11.171

Analysis of Deviance Table

Model 1: infected ~ weight
Model 2: infected ~ weight + sex
Resid. Df Resid. Dev Df Deviance
1        79     68.072
2        78     63.785  1    4.287
```

(e) What would you conclude from the output above?

[~25% marks]

(f) Provide one or two sentences that you could use in a paper to summarise your analysis.

[~15% marks]

SECTION C

C5 The giant hogweed, *Heracleum mantegazzianum*, was introduced into Great Britain by the Victorians, and has been spreading ever since. The model

$$\frac{dN}{dt} = \big(b(N) - d(N)\big)N,$$

is suggested, where $b(N)$ and $d(N)$ are functions of the size of the hogweed population at time t, $N(t)$. A particular version of the model has

$$b(N) = b_0 - b_1 N,$$
$$d(N) = d_0,$$

in which $b_0 > d_0$.

(a) Interpret the biological meaning of the model. *[~10% marks]*

(b) Demonstrate that if β and κ are chosen appropriately, the model can be written as

$$\frac{dN}{dt} = \beta N\left(1 - \frac{N}{K}\right).$$

Interpret the meaning of the expressions for β and K.

[~20% marks]

(c) Sketch the direction field associated with the model, and use it to sketch $Y(t)$, assuming that the initial hogweed population is small.

[~20% marks]

An alternative model takes a different form for $b(N)$

$$b(N) = \frac{b_0}{1 + b_1 N},$$
$$d(N) = d_0.$$

(d) Sketch $b(N)$ and interpret the biological basis of the updated model, suggesting one reason why it might be an improvement.

[~15% marks]

(e) Determine the population size at which the total number of Hogweeds is increasing most quickly according to the updated model.

[~20% marks]

(f) Suggest three features of even the updated model that may be unrealistic for an invading plant species. *[~15% marks]*

C6 A group of organisms that is living in an isolated habitat has population size at time t, $Y(t)$. The population is subject to immigration at rate α, reproduction at per capita rate β, and death at per capita rate γ, where α, β and γ are all positive constants. In parts (a)-(d) you should assume that $\beta < \gamma$.

(a) Write down a model of the form

$$\frac{dY}{dt} = F(Y),$$

to describe the evolution of $Y(t)$. *[~10% marks]*

(b) Find the equilibrium value of your model, and examine its stability using a method based on calculating dF/dY at the model's equilbrium.

[~25% marks]

(c) Solve the model to find $Y(t)$, given that the habitat was initially empty. Verify your solution is consistent with your answer to part (b).

[~30% marks]

(d) Sketch your solution and interpret the changes to your sketches as the values of the parameters are changed. *[~20% marks]*

(e) What would happen if instead parameters were such that $\beta = \gamma$?

[~15% marks]

SECTION D

D7 The dynamics of a system are described by the following pair of simultaneous first-order non-linear differential equations.

$$\frac{dx}{dt} = xy + x^2,$$
$$\frac{dy}{dt} = 6 - x^2 - y.$$

(a) Find the equilibrium point(s). *[~20% marks]*

(b) Classify the equilibrium point(s). *[~30% marks]*

(c) Sketch the null-clines for this system, and then the phase plane, showing the behaviour around the equilibrium points by adding representative trajectories. *[~30% marks]*

(d) For the initial condition (-2, 0) sketch the trajectory of the path on the phase plane, and sketch the graph of x against t, showing the main qualitative features. *[~20% marks]*

D8 Suppose that a dangerous new virus of humans has emerged and has circulated in several countries other than the United Kingdom. Last week the first cases were found in London. As far as we know the disease is not fatal and people recover with long lasting immunity. The following equations describe the dynamics of this viral infectious disease.

$$\frac{dS}{dt} = bN - bS - \beta IS,$$
$$\frac{dI}{dt} = \beta IS - \nu I - bI,$$
$$\frac{dR}{dt} = \nu I - bR.$$

(a) Briefly give biological definitions of each of the model parameters.
[~15% marks]

(b) Define R_0 and derive an expression for this quantity for the model above.
[~20% marks]

(c) Find the non-zero equilibrium. *[~20% marks]*

Fortunately a vaccine has already been developed, and this vaccine confers perfect immunity against the virus.

(d) How would R_0 and the non-zero equilibrium point have been altered if a proportion p of the population had been instantaneously vaccinated at $t = 0$?
[~20% marks]

It is discovered that newborn children are most at risk of severe disease following infection.

(e) Rewrite the model such that, instead of instantaneously vaccinating a proportion p of the population at $t = 0$, instead a proportion q of newborns is vaccinated routinely at the time of birth. What is the minimum proportion that needs to be vaccinated to prevent an outbreak?
[~25% marks]

SECTION E

ANSWER NO MORE THAN 2 QUESTIONS FROM SECTION E

E9 Two lionesses are confronted by a potential prey of energetic value b. Each must simultaneously decide whether or not to pursue it. Pursuit entails the expenditure of c units of energy. If only one lioness pursues the prey, she captures it with probability p_1, in which case she consumes it all and thus gains b units of energy (while the other lioness gains nothing). If both lionesses pursue the prey, they capture it with probability p_2 ($> p_1$), in which case they share it equally and thus gain $b/2$ units of energy each. If neither lioness pursues the prey, neither gains anything.

(a) Write down a payoff matrix for this game (with payoffs in terms of net expected energetic gain). *[~10% marks]*

(b) Under what conditions is each of the pure strategies in this game evolutionarily stable? *[~20% marks]*

(c) Under what conditions does the game yield an evolutionarily stable mixed strategy? Derive an expression for the evolutionarily stable probability of pursuit under these conditions.

[~30% marks]

(d) Suppose that the two lionesses represent a mother and her daughter, and that if both pursue and capture the prey, the mother will then claim the greater share, gaining $2b/3$ units of energy (while the daughter gains only $b/3$). Write out a new payoff matrix for the game, and determine the conditions under which the strategy “pursue if mother, but not if daughter” is evolutionarily stable.

[~40% marks]

ANSWER NO MORE THAN 2 QUESTIONS FROM SECTION E

E10

(a) For an enzyme catalysed reaction, show how the steady state approximation and other assumptions can be used to derive the Michaelis-Menten equation:

$$v_0 = \frac{V_{max}[S]}{K_M + [S]}$$

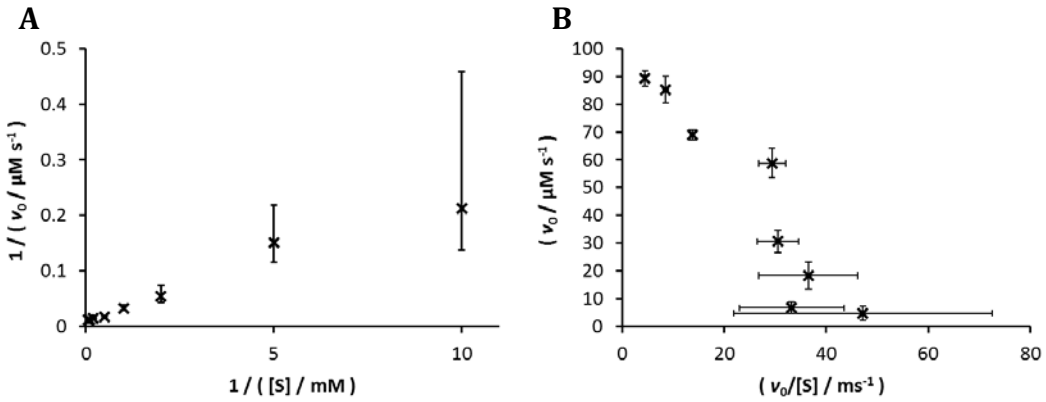
where v_0 is the initial rate of product formation, **[S]** is the initial concentration of the substrate S, and V_{max} and K_M are positive constants.

Base your answer on the following reaction scheme, where E represents the enzyme, (ES) is the enzyme-substrate complex, P is the product and k_1, k_{-1} and k_2 are rate constants for the indicated steps:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} (ES) \xrightarrow{k_2} P + E$$

[~60% marks]

(b) Panels **A** and **B** below show scatter plots for a data set of substrate concentrations **[S]** and initial rates v_0 for the same enzyme catalysed reaction after attempting to apply two different linear transformations. For each case, explain why it would not be appropriate to perform an unweighted linear regression using the horizontal axis coordinate as the explanatory variable and the vertical axis coordinate as the response variable.



[~20% marks]

(c) If the data set discussed in part (b) is expected to obey the Michaelis-Menten equation, suggest an alternative approach that would treat the experimental data more appropriately and would yield accurate values for the fitting parameters V_{max} and K_M.

[~20% marks]

ANSWER NO MORE THAN 2 QUESTIONS FROM SECTION E

E11

(a) The following dynamic programming sequence alignment matrix was completed using scores from an amino acid substitution matrix; a fixed penalty was used for gaps.

	-	T	L	W	V	N	K	C	H	V	Q
-	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	5	0	0	0	1
W	0	0	0	11	3	0	0	3	0	0	0
A	0	0	0	3	11	3	0	0	1	0	0
E	0	0	0	0	3	11	4	0	0	0	2
N	0	0	0	0	0	9	11	3	1	0	0
R	0	0	0	0	0	1	11	8	3	0	1
C	0	0	0	0	0	0	3	20	12	4	0
W	0	0	0	11	3	0	0	12	18	10	2

(i) Was this matrix completed for finding a global or local alignment? How can you tell?

[~5% marks]

(ii) What is the gap penalty?

[~5% marks]

(iii) Show the highest scoring local alignment between the two sequences and the score at each position in the alignment.

Hint. Follow the format of the following example to illustrate an alignment and the score at each position:

```
            P  P  E  G  R  H
            |  |           |
            P  P  D  -  K  H
scores:     7 14 16  8 10 18
```

[~10% marks]

(QUESTION E11 CONTINUES OVERLEAF; TURN OVER)

E11 (continued)

(b)

(i) Copy and complete the dynamic programming sequence alignment matrix given below for finding an optimal global alignment of the sequences ATTG and GACT.

Use the following scoring scheme: nucleotide match = +3, nucleotide mismatch = −1, gap penalty = −1. Use arrows to show the potential trace back options for *every* cell.

	-	**A**	**T**	**T**	**G**
-	**0**				
G					
A					
C					
T					

[~25% marks]

(ii) What is the score of the optimal global alignment?

[~5% marks]

(iii) Mark (e.g. with circles) the optimal global alignment path on the matrix.

[~5% marks]

(iv) Show the optimal global alignment between the two sequences and the score at each position in the alignment.

Hint. Follow the format of the following example to illustrate an alignment and the score at each position:

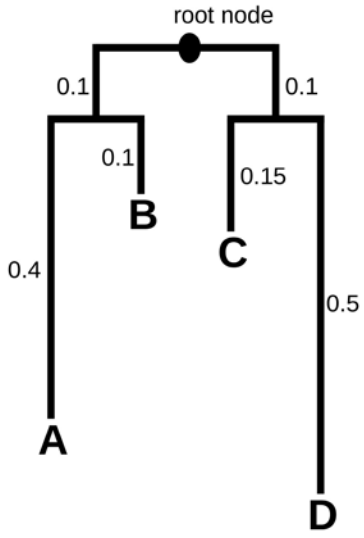
```
          C G G U A G C
          |   |   |   |
          C A G - A U C
scores:   3 2 5 4 7 6 9
```

[~10% marks]

(QUESTION E11 CONTINUES ON THE NEXT PAGE)

E11 (continued)

(c) In the rooted phylogenetic tree below, vertical branch lengths (with numbers) correspond to evolutionary distance. The root represents an ancestral sequence whereas A, B, C and D represent four present-day sequences.



(i) Does this tree have a constant molecular clock? How can you tell?

[~5% marks]

(ii) Calculate the values of *w*, *x*, *y* and *z* in the distance matrix below, in which the distance measure is evolutionary distance.

	A	**B**	**C**	**D**
A	**0**			
B	***w***	**0**		
C	***x***	**0.45**	**0**	
D	***y***	***z***	**0.65**	**0**

[~10% marks]

(iii) The UPGMA algorithm for constructing a phylogenetic tree assumes that the tree has a constant molecular clock. However, it is often still possible to apply the UPGMA algorithm even when this condition is not satisfied, although the resulting tree may be different from the true evolutionary tree. Apply the UPGMA algorithm to the distance matrix from part ii. Show the tree produced by the UPGMA algorithm and annotate the branch lengths according to the UPGMA algorithm (note that the resulting tree branch lengths may disagree with the distance matrix from part ii).

[~20% marks]

END OF PAPER