apts.ac.uk
Academy for PhD Training in Statistics

APTS Module Specifications

APTS Executive Committee
with help from the APTS Module Leaders

Revised 2007-06-27

1 Introduction

The APTS programme is for first-year PhD students in statistics and applied probability. The programme comprises 8 taught-course modules, to be delivered intensively in four pairs. This document concerns the specification of the 8 modules.

The aims of APTS modules differ from those of typical undergraduate-level or MSc-level lecture courses. Broadly, the aims are to equip first-year doctoral students with a clear 'mental map' of some areas of work that are important in current statistical research, and to help them gain confidence and resources (e.g., analytical tools, computing tools, literature entry-points) to find out more for themselves, as and when needed. The backgrounds of APTS students will vary widely, and it will be crucial to make APTS material accessible to all of those taking part; the APTS modules are intended primarily as *preparation* for more advanced study, and they should not themselves have ambitions which are too advanced.

We have aimed at a consistent format:

(a) **aims, learning outcomes and prerequisites** to specify the core of the module;

(b) an **indicative list of topics**;

(c) an indication of the sort of **assessment** that might be offered.

Here are more comments about each of these in turn.

(a) Aims, learning outcomes and prerequisites

The aims, learning outcomes and prerequisites are written to help specify how each module fits in with the other 7 APTS modules, and to help achieve a general similarity of level. In particular it should be noted:

(i) Module teaching time will be limited. The time allocated to each of the two modules delivered in a 5-day 'APTS week' is 10.5 hours in total. The precise balance within this — between lectures, classes, practicals etc. — is at the discretion of module leaders. A suggested norm is 1.5 hours of lecture per day on each module, with the remaining time being devoted to practical/exercise classes.

(ii) Students' prior knowledge will be uneven (APTS will draw from a wide range of UK universities and student backgrounds). Associated with each module will be a couple of preliminary web-lectures (format to be determined: e.g., some module leaders may want to use video etc, others will not), and module leaders will supply relevant material to help students get up to speed; but over-loading students with such material should be avoided. We have tried therefore to keep aims and learning outcomes at a manageably small size. The 'prerequisites' stated in the module specifications below are of three main types:

  (a) background expected from undergraduate or MSc education, or from preliminary PhD training in students' home departments;

  (b) topics from other APTS modules;

(c) other preliminary material specific to the module concerned.

The preliminary web-lectures for each module will review briefly what is needed from students under (a) and (b), and will provide students with, or point them towards, self-study materials (lecture notes, exercises, reading) to enable them to get to grips with (c). The amount of work that students will need to do themselves prior to an APTS week will vary widely, depending for example on the courses they have taken as undergraduates. In the spirit of the APTS aims as stated in the introductory paragraphs, in some cases the best plan may be for a student to take an undergraduate or MSc-level course, or to undertake a course of directed reading. (An example of this might be a student who has not met the notions of Markov chain or autoregression, and who will later need to acquire expertise for their research.) The preliminary web-lectures and associated material will provide such students with a minimal acquaintance with the topic, sufficient for the student to acquire the relevant mental map through participation in the module, bearing in mind that supplementary courses/reading may be undertaken *after* the APTS module. Module leaders are encouraged to exercise their own judgement in determining what can reasonably be expected of students prior to the APTS week, and which of the 'prerequisites' will need to be reinforced during the APTS week itself.

(b) List of topics

Given the above, these lists are intended to indicate possibilities rather than to stipulate requirements. The ethos of APTS modules is:

- to provide tools and material to gain basic understanding of the relevant area;
- to indicate difficulties and sketch solutions;
- to point authoritatively to the relevant literature;
- to emphasise applications, wherever relevant;
- to give a taste for further study.

Our subject is huge and growing rapidly; it will be of great value for students to acquire perspectives by listening to authoritative experts on these various subjects. For example, a module leader may choose to give a brief overview of what is involved in the area, and select a few examples or special cases to discuss properly rather than attempt an account which aspires to be encyclopaedic but ends up merely being indigestible.

(c) Assessment

APTS is resourced (not lavishly) to supply PhD training through APTS modules, but has not been resourced at all to conduct assessment of students. Module leaders are therefore asked only to supply material for light-touch assessment, which should be in a form useable by a generalist in a student's home department to gain a modest appraisal of student learning via the module. The assessment tasks provide the opportunity for each department to fit APTS training into its own framework for monitoring student progress.

In the module specifications below, the sentences under this heading specify examples varying per module, mainly in order to communicate a sense of what would suffice. Ideally the assessment task should be constructed in such a way that it adds to what has been learned by students during the APTS intensive week. It should enhance both the students' interest in the area of the module and their ability to find out more for themselves.

2 The module specifications

The eight APTS modules are:

Statistical Computing	Statistical Inference
Statistical Modelling	Statistical Asymptotics
Applied Stochastic Processes	Computationally-intensive Statistical Methods
Spatial and Longitudinal Data Analysis	Nonparametric Smoothing

Statistical Computing

Module leader: S. N. Wood

Aim: To introduce, in a practical way, the fundamentals of numerical computation for statistics, in order to help students to write stable, fast and numerically accurate statistical programs.

Learning outcomes: After taking this module students will:

- understand the importance of stability, efficiency and accuracy in numerical computations, and how these may be promoted in practical statistical computation;
- understand the main difficulties and other issues that arise in the topics given below;
- be aware of standard computational libraries and other resources.

Prerequisites: In preparation for this module, students should obtain an elementary knowledge of the use of R. (Knowledge also of a lower level language such as C, Pascal or Fortran would be an advantage but will not be presumed.) Preparation for this module should also (re-)establish familiarity with Taylor's theorem and with basic matrix algebra — e.g., notion of an inverse and eigenvalues, manipulation of matrix expressions, the numerical unsuitability of Cramer's rule for computation of an inverse.

Topics:

- Finite-precision arithmetic; related types of error and stability (probably mostly covered, in context, as part of other topics).
- Numerical linear algebra (with statistical applications): basic computational efficiency, Choleski, QR, stability (e.g. Normal/Choleski vs QR for LS), eigen and singular value decompositions. Standard libraries.
- Optimization: Newton-type methods; other deterministic methods; stochastic methods; using methods effectively in practice; what to use when.
- Differentiation and integration by computer: finite differencing (interval choice, cancellation and truncation errors); automatic differentiation; quadrature methods; stochastic integration.
- Basics of stochastic simulation.
- Other types of problem (e.g. sorting and matching); the pervasiveness of efficiency and stability issues; where to find out more.

Assessment: A short project bringing together several of the topics covered. For example writing a routine to estimate a linear mixed model by (RE)ML.

Statistical Inference

Module leaders: D. R. Cox and D. Firth

Aims: This module will provide students with a solid understanding of the main approaches to statistical inference, their strengths and limitations, their similarities and differences, and their role in underpinning statistical methodology.

Learning outcomes: After taking this module students should have an appreciation of the predominant modes of inference and their inter-relationships, and should be better equipped to read the published literature on both technical and foundational aspects of inference.

Prerequisites: Students should review the following definitions and results: likelihood, sufficiency, Bayes' theorem; simple properties of normal, exponential, binomial and Poisson distributions; linear model and the method of least squares.

Topics:

- Role of formal inference, nature of probability, frequentist and Bayesian approaches.
- Role of sufficiency; role of Neyman-Pearson theory; relation between significance tests and confidence limits.
- Maximum likelihood and associated issues; properties in 'standard' situations, and in some more difficult cases.
- Exponential-family models.
- Other approaches (e.g., estimating equations, pseudo-likelihoods).

Assessment – one of:

- An essay on one of a list of topics suggested by the module leaders.
- Report of a numerical investigation on one of a list of topics suggested by the module leaders, e.g., comparing Bayesian and frequentist approaches to the analysis of a particular model, or assessing the accuracy of inferences based on large-sample approximation.

Statistical Modelling

Module leaders: A. C. Davison and J. J. Forster

Aims: The main aim of this module is to introduce important general aspects of statistical modelling, including Bayesian modelling. A broad range of specific, commonly-used types of model will also be encountered.

Learning outcomes: After taking this module, students should — for topics listed below which are included in the module — understand the issues (why this is important), the terminology, the statistical principles associated with this aspect of modelling, and sufficient theory to deal with simple examples; and they will have gained some practical hands-on experience in more complex examples.

Prerequisites: Preparation for this module should (re-)establish familiarity with linear and generalized linear models, and with likelihood and Bayesian inference. Students who are familiar with (for example) chapters 4, 8, 10 and 11 of Davison's *Statistical Models* will be very well prepared (and will already know something of the areas to be covered in the module).

Topics:

- Missing data and latent variables.
- Principles and practice of model selection.
- Random-effects/hierarchical/mixed models.
- Semiparametric models and smoothing (links with the later APTS module 'Nonparametric Smoothing').
- The role of conditional independence in modelling. Introduction to graphical models.

Assessment: Either a suitably constructed 'comprehension exercise', for which students read a recent paper from the literature involving advanced modelling, and answer a series of questions, some of which may be quite open-ended; or a practical exercise involving real data and research questions.

Statistical Asymptotics

Module leader: G. A. Young

Aims: This module has the twin aims of introducing students to asymptotic theory and developing their practical skills in using asymptotic approximations.

Learning outcomes: After taking this module, students will have a basic understanding of the asymptotic properties of parametric likelihoods and posterior distributions, and the knowledge and skills to derive and implement first-order Laplace and saddlepoint density approximations in simple examples.

Prerequisites: Preparation for this module should establish:

- basic knowledge of likelihood methods, exponential families and Bayesian inference, to the level developed in a typical third-year undergraduate inference course;
- knowledge of limit theorems in the univariate IID case (laws of large numbers and CLT);
- familiarity with different modes of convergence (convergence in distribution, in probability, almost sure and L^p);
- familiarity with Taylor expansions in the multivariable case;
- familiarity with $o(.)$, $O(.)$, $o_P(.)$ and $O_P(.)$ notation.

Topics:

- Multivariate central limit theorem, (a gentle introduction to) the continuous mapping theorem, the delta method;
- Stochastic asymptotic expansion;
- Likelihood asymptotics (including asymptotic properties of MLEs);
- Asymptotic normality of posterior distributions (parametric case);
- Laplace's approximation (univariate and multivariate);
- Introduction to Edgeworth expansions and saddlepoint density approximations (via tilting);
- Saddlepoint approximations to tail probabilities.

Assessment: A mini-project which ideally has both a theoretical component (e.g., discussion of conditions for asymptotic normality in a particular set-up, or derivation of a suitable approximation in particular examples) and a computational component (e.g., numerical implementation of a Laplace or saddlepoint approximation).

Applied Stochastic Processes

Module leader: W. S. Kendall

Aims: This module will introduce students to two important notions in stochastic processes — reversibility and martingales — identifying the basic ideas, outlining the main results and giving a flavour of some of the important ways in which these notions are used in statistics.

Learning outcomes: A student successfully completing this module will be able to:

- describe and calculate with the notion of a reversible Markov chain, both in discrete and continuous time;

- describe the basic properties of discrete-parameter martingales and check whether the martingale property holds;
- recall and apply significant concepts from martingale theory (indicative list: optional stopping, martingale convergence);
- explain how to use Foster-Lyapunov criteria to establish recurrence and speed of convergence to equilibrium for Markov chains.

Prerequisites: Preparation for this module should include a review of the basic theory and concepts of Markov chains as examples of simple stochastic processes (transition and rate matrices, irreducibility and aperiodicity, equilibrium equations and results on convergence to equilibrium), and with the definition and basic properties of the Poisson process (as an example of a simple counting process).

Topics:

- Reversibility of Markov chains in both discrete and continuous time, computation of equilibrium distributions for such chains, application to important examples.
- Discrete time martingales, examples, application, super-martingales, sub-martingales.
- Stopping times, statements and applications of optional stopping theorem, martingale convergence theorem.
- Recurrence and rates of convergence for Markov chains, application to important examples.
- Statements and applications of Foster-Lyapunov criteria, viewed using the language of martingales.
- Statistical applications and relevance (highlighted where appropriate throughout).

Assessment: One of

- Read an appropriately chosen paper (either specified or from a specified list), and identify some of the main stochastic process models and results referred to in the paper, describe their significance and use in the paper, and relate them to the material covered in the module.
- Complete appropriate exercises that are simple developments or extensions of aspects of the results in the module (that may be related in some way to models and results arising in the paper).

Computationally-intensive Statistical Methods

Module leader: B. D. Ripley

Aims: This module will introduce various computationally-intensive methods and their background theory, including material on simulation-based approaches such as Markov-chain Monte Carlo (MCMC) and the bootstrap, and on strategies for handling large datasets. The different methods will be illustrated by applications.

Learning outcomes: After taking this module, students will have a working appreciation of MCMC, the bootstrap and other simulation-based methods and of their limitations, and have some experience of implementing them for simple examples. Students will also have gained an appreciation of the difficulties of handling very large datasets and of some approaches to overcoming them.

Prerequisites: Preparation for this module should include a review of:

- relevant basic material on statistical modelling (for which the earlier APTS module 'Statistical Modelling' would be advantageous);
- basic Markov chains (as for the 'Applied Stochastic Processes' module).

Basic knowledge of programming in a high-level language such as *R* will be assumed, and *R* will be used for case studies and exercises.

Topics:
- Overview of simulation-based inference; Monte Carlo testing.
- Basic theory of bootstrap methods; practical considerations; limitations.
- Basic theory of MCMC; types of MCMC samplers; assessment of convergence/mixing; other practical considerations; case studies.
- Strategies for dealing with large datasets: use of database management systems, multipass algorithms, subsampling, distributed computing. A case study, e.g. generalized linear models.

Assessment: Analysis of a data set using computationally-intensive methods. This could be a data set related to the student's research project, or an example specified by the module leader.

Spatial and Longitudinal Data Analysis

Module leader: P. J. Diggle

Aims: This module will introduce students to the statistical concepts and tools involved in modelling data which are correlated in time and/or space. The content will include models which are well established in statistical practice, although not usually well represented in the undergraduate curriculum, as well as examples of models which are central to current research in the area.

Learning outcomes: By the end of the module, students should have achieved:

- a clear understanding of the meaning of temporal and spatial correlation;
- a good working knowledge of standard models to describe both the systematic and the random parts of an appropriate model;
- the ability to implement and interpret these models in standard applications;
- an understanding of some of the key concepts which lie at the heart of current research in this area;
- appreciation of at least one substantial case study.

Prerequisites: Preparation for this module should establish familiarity with:

- standard models and tools for time series data, at the level of a typical undergraduate course on time series;
- standard models and tools for spatial data at its simplest level;
- inferential methods, including classical and Bayesian likelihood-based methods, to at least the level of the earlier APTS modules 'Statistical Inference' and 'Statistical Modelling'.

This module's preliminary web-lectures will cover the first two of the above pre-requisites.

Topics:
- Introduction: motivating examples; the fundamental problem — analysing dependent data.
- Longitudinal data: linear Gaussian models; conditional and marginal models; why longitudinal and time series data are not the same thing.
- Continuous spatial variation: stationary Gaussian processes; variogram estimation – what not to do and how to do it; likelihood-based estimation; spatial prediction.
- Discrete spatial variation: Markov random field models.
- Spatial point patterns: exploratory analysis; Cox processes and the link to continuous spatial variation; pairwise interaction processes and the link to discrete spatial variation.
- Spatio-temporal modelling: spatial time series; spatio-temporal point processes.
- Conclusion: review of available software (as preparation for mini-project); connections — spatial and longitudinal data analysis as two sides of the same coin.

Assessment: One of

- A critique, in essay form, of a specified research paper, including both modelling and application aspects;
- A mini-project involving the analysis of a data-set, selected by the student from several on offer (to allow students to focus on topics within the course which they find particularly interesting).

Nonparametric Smoothing

Module leader: A. Delaigle

Aims: In nonparametric statistics, it is not assumed that data come from a distribution belonging to a finite-dimensional class. This module will introduce students to the theory and methods associated with the idea of smoothing, which is one of the key concepts in modern nonparametric techniques of statistical analysis.

Learning outcomes: By the end of the module, students will understand the technique of kernel density estimation and its advantages over histograms. They will appreciate the central role played by the smoothing parameter, or bandwidth, and understand how this can be determined in practice. Students will also understand how both kernel- and spline-based smoothing methods can be used to estimate a nonparametric regression function.

Prerequisites: Preparation for this module should include a rapid review of basic probabilistic limit theory (including modes of convergence, laws of large numbers, 'o'and 'O' notation and the delta method), and basic understanding of asymptotic expansions — to at least the level of the earlier APTS module 'Statistical Asymptotics'.

Topics:

- Kernel and spline approaches to smoothing;
- Determination of degree of smoothing (bandwidth, penalty, effective degrees of freedom);
- Density estimation;
- Nonparametric regression;
- Applications, e.g., covariate measurement error, generalized additive models.
- Connections with wavelets and other nonparamatric estimators.

Assessment: A set of exercises assigned by the module leader, or a data-analysis exercise involving practical use of some of the methods covered.