Postgraduate Statistics Centre
Department of Mathematics and Statistics

LANCASTER UNIVERSITY

Masters in Data Science: Statistical Inference Specialism

Module Outlines

MSc Data Science: Statistical Inference Specialism

1. Course Overview

The Statistical Inference specialism is aimed at students with a background in mathematics and statistics and who want to develop their statistical, computing and analytical skills for the extraction, synthesis, processing and analysis of large and complex data. It is a part of a multi-disciplinary degree programme shared across the Faculty of Science and Technology: Mathematics and Statistics; the School of Computing and Communications and Lancaster Environment Centre. Building upon university specialism and expertise a number of the optional modules will also be contributed from Lancaster University Management School.

The programme will encompass data science fundamentals but will have added focus upon statistical modelling and inference of large and complex data structures. More specifically, the course will provide a thorough training in statistical theory; data analysis and computing via a distinctive blend of leading-edge methodology and practical techniques including Bayesian methods, machine learning, data mining and forecasting. More generally, optional modules spanning, for example, genomics, longitudinal data analyses, time-to-event data and spatial data analyses will be available for study. The modules reflect inter-departmental research expertise and prepare students for particular career options in areas with growing demand for data scientists.

The Statistical Inference specialism of MSc in Data Science consists of a series of **taught modules** followed by the completion a **dissertation**.

The taught course component consists of 9 modules which can be decomposed as follows:

- A 'core' set of five compulsory modules spanning modern data science including: computing, data synthesis and extraction, visualisation, data mining, statistical modelling and likelihood inference;
- A `compulsory specialist module' in Bayesian methods;
- A set of three `optional' modules (chosen from 10) spanning a range of specialist/advanced statistical methods relevant to the design, analysis and interpretation of observational and experimental data.

The dissertation component consists of an in-depth project applying data scientific methods to address substantive research question. This will be undertaken in the summer term and will typically incorporate a two-month internship.

Statistical Inference Course Structure

A 'core' set of modules in common with the Computing specialism and the Data Science for the Environment programme. The mathematical content of the statistics modules are tailored according to background

Module	Title	Coursework	Exam %
SCC450	Data Science Fundamentals	100	NA
SCC403	Data Mining	100	NA
SCC446	Programming for Data Scientists	100	NA
MATH551	Likelihood inference	50	50
MATH552	Generalised linear models	50	50

A 'compulsory' Bayesian Inference module

Module	Title	Coursework	Exam %
MATH553	Bayesian inference	50	50

Three 'self-selected' optional modules from:

Module	Title	Coursework	Exam %
MATH562	Extreme value theory	50	50
MATH563	Clinical trials	50	50
MATH564	Principles of epidemiology	50	50
CHIC565	Environmental epidemiology	50	50
MATH566	Longitudinal data analysis	50	50
MATH572	Genomics: technologies and data analysis	100	NA
MATH573	Survival and event history analysis	50	50
MSCI523	Forecasting	100	NA
MSCI526	Data Mining for Marketing, Sales and Finance	100	NA
MSCI534	Optimisation & Heuristics	30	70

SCC modules taught by School of Computing and Communications; MATH modules: taught by the Department of Mathematics and Statistics; MSCI modules taught by Lancaster University Management School: CHIC taught by CHICAS research group.

MODULE DESCRIPTIONS

SCC 450 Data Science Fundamentals

Duration: 20 Weeks (25 learning hours) over Michaelmas and Lent Term

Credits: 15

Department: SCC/M&S/LEC

This module will provide a common core to the entire degree programme and will run throughout the Michaelmas and Lent terms.

This module will provide a common core to the entire degree programme and will run throughout the Michaelmas and Lent terms. Within this module students will learn about the Data Science pipeline, as presented above, and how the combination of applied research, mathematics and statistics, and computing, all contribute to various steps along the way. Students will be given lectures about understanding research problems and formulating research questions from different disciplines (Environmental Science, Computer Science, and Maths and Stats) in order to ground their understanding of how research is conducted in different fields. They will work together on research problems and present their findings in a non-technical manner, thereby replicating the industrial setting of conveying technical information to a non-technical audience. Students will learn how to write up and report their findings in a manner that can be acted upon by their colleagues within an industrial setting. Students will learn about ethics and research design that considers this issue within the Data Science process, when data should and shouldn't be collected, and what an employee's obligations are for protecting the data and for performing inferences over the data. This module will be taught between departments, thereby drawing on research and industrial experience from SCC, M&S and LEC.

Recommended Reading:

Doing Data Science: Straight Talk from the Frontline. 1st edition. Cathy O'Neil, and Rachel Schutt. November 2013, ISBN 978-1449358655.

SCC403 Data Mining

The existing SCC module on Data Mining will be a core offering to students taking any Data Science MSc programmes offered by LEC, M&S and SCC.

The course content includes aspects related to: data representation and storage and the different paradigms that these follow such as vector-space models, relational databases and NoSQL databases. Data processing techniques will be taught including data mining and machine learning, such as clustering methods for grouping data, classification techniques (e.g. Support Vector Machines), and fuzzy-classifiers. Learning routines will also be taught such as semi-supervised learning through self-training.

Recommended Reading:

Data Mining: Practical Machine Learning Tools and Techniques . 3rd edition. Ian H. Witten, Mark A. Hall, and Elbe Frank. February 2011, ISBN 978-0123748560.

SCC446 Programming for Data Scientists

This module on Programming for Data Scientists will be a core offering to students taking any Data Science MSc programmes offered by LEC, M&S and SCC.

This module will cover statistical programming using R, and object-oriented programming using Java. The former language forms a core component of statistical analysis and is therefore essential for analysing and interpreting data. The latter language is essential for processing and integrating data from multiple disparate sources, and for dealing with data heterogeneity. M&S will teach the first 5 weeks of this module on programming with R, while SCC will teach the second 5 weeks of this module on object-oriented programming using Java. Assessment will be through logbook exercises throughout the course.

MATH551 Likelihood Inference

Prerequisites: UG Mathematics/Statistics (probability theory; calculus; matrices etc)

Description:

Statistical theory is the theory of the extracting information about the unknown parameters of an underlying probability model from observed data. This underpins all practical statistical applications, such as those considered in later MSc modules.

This course considers the idea of statistical models, and how the likelihood function, the probability of the observed data viewed as a function of unknown parameters, can be used to make inference about those parameters. This inference includes both estimates of the values of these parameters, and measures of the uncertainty surrounding these estimates. We consider single and multi-parameter models, and models which do not assume the data are independent and identically distributed. We also cover basic computational aspects of likelihood inference that are required in many practical applications.

Objectives

On successful completion students will:

understand how to construct statistical models for simple applications; appreciate how information about the unknown parameters is obtained and summarized via the likelihood function; be able to calculate the likelihood function for independent and identically distributed data; be able to calculate the likelihood function for some statistical models which do not assume independent identically distributed data; be able to evaluate point estimates and make statements about the variability of these estimates; understand about the inter-relationships between parameters, and the concept of orthogonality; be able to perform hypothesis tests using the generalised likelihood ratio statistic; use computational methods to calculate maximum likelihood estimates; use computational methods to construct confidence intervals, and perform hypothesis tests; be able to look at residuals to judge how appropriate a model is.

Syllabus:

The course presents the key tools for statistical inference, stressing the fundamental role of the likelihood function. It will cover:

- Definition of the likelihood function for single and multi-parameter models, and how it is used to calculate point estimates (maximum likelihood estimates)
- Asymptotic distribution of the maximum likelihood estimator, and the profile deviance, and how these are used to quantify uncertainty in estimates
- Inter-relationships between parameters, and the definition and use of orthogonality
- Generalised Likelihood Ratio Statistics, and their use for hypothesis tests
- Calculating likelihood functions for non-iid data
- Simple use of computational methods to calculate maximum likelihood estimates and confidence intervals and to perform hypothesis tests
- Model criticism through residual analysis

Bibliography:

Azzalini Statistical Inference: Based on the Likelihood. Chapman and Hall 1996

D R Cox and D V Hinkley Theoretical Statistics. Chapman and Hall 1974

Y Pawitan In All Likelihood: Statistical Modeling and Inference Using Likelihood. OUP 2001

MATH552: Generalised Linear Models

co-requisite: MATH551.

Objectives of the module:

Generalised linear models are now one of the most frequently used statistical tools of the applied statistician. They extend the ideas of regression analysis to a wider class of problems that involves exploring the relationship between a response and one or more explanatory variables. In this course we aim to discuss applications of the generalised linear models to diverse range of practical problems involving data from the area of biology, social sciences and time series to name a few and to explore the theoretical basis of these models. By the end of the course students should be able to formulate sensible models for sets of data, taking account of the motivation for data collection; fit and check these models in the statistical package R; produce confidence intervals and tests corresponding to questions of interest; and state conclusions in everyday language.

Description:

We introduce a large family of models, called the generalised linear models (GLMs), that includes the standard linear regression model as a special case and to discuss the theoretical properties of these models.

We learn a common algorithm called iteratively reweighted least squares algorithm for the estimation of parameters.

We fit and check these models with the statistical package R; produce confidence intervals and tests corresponding to questions of interest; and state conclusions in everyday language.

Syllabus:

- **Model specification**: choosing a suitable GLM; equivalent models; aliasing.
- **Fitting models**: maximum likelihood estimation using R.
- **Effects of covariates**: confidence intervals and tests of quantities of interest, interaction.
- **Variable selection:** backwards stepwise selection of covariates.
- **Assessing model fit**: deviance and Pearson residuals; leverage; residual deviance test of model fit; over-dispersion.

Bibliography:

P. McCullagh and J. Nelder. Generalized Linear Models, Chapman and Hall, 1999.

A.J. Dobson, An Introduction to Generalised Linear Models, Chapman and Hall, 1990.

MATH553: Bayesian Inference

Prerequisites: MATH551; MATH552

Objectives of the module:

This course aims to introduce the Bayesian view of statistics, stressing its philosophical contrasts with classical statistics, its facility for including information other the data into the analysis and its coherent approach toward inference and model section. By the end of the course the students should be able to formulate an appropriate prior to a variety of problems, calculate, simulate from and interpret the posterior and the predictive distribution, to carry out Bayesian model selection using the marginal likelihood and model checking using the predictive distribution. Students should be able to carry out all of this using the programming language R.

Syllabus:

- Inference by updating belief
- The ingredients of Bayesian inference: the prior, the likelihood, the posterior, the predictive and the marginal distributions.
- Methods for formulating the prior
- Conjugate priors for single parameter models
- Regression
- Sampling from the posterior and the predictive distribution in the case of conjugate priors.
- For non-conjugate priors, sampling from the posterior using an approximation and then correcting for the approximation using importance sampling-resampling methods.
- Rejection sampling
- Methods for summarising the posterior distribution Model checking using the predictive and model selection using the marginal likelihood

Bibliography:

Gamerman and Lopez (2006). MCMC statistical simulation for Bayesian inference. Chapman and Hall 2nd Edition

Hoff, P (2008). A first course in Bayesian statistics. Springer.

Petris et al., (2009), Dynamic Linear Models with R, Springer UseR series

MATH562 Extreme Value Theory

Prerequisites: MATH551; MATH552

Objectives of the module:

This module aims to develop the asymptotic theory, and associated techniques for modelling and inference, associated with the analysis of extreme values of random processes. The course will focus on the mathematical basis of the models, the statistical principles for implementation and the computational aspects of data modelling. Students are expected to acquire the following: an appreciation of, and facility in, the various asymptotic arguments and models; an ability to fit appropriate models to data using specially developed R software; the ability to understand and interpret fitted models.

Description:

For many physical processes, especially environmental processes, it is extremes of the process that are of greatest concern; the highest sea-levels cause floods; the fastest wind-speeds destroy buildings, etc. Most of the statistical theory is concerned with modelling typical behaviour; in contrast, the analysis of extremes requires us to model the unusual. This means that we have very little data with which we can either develop or estimate models. In the absence of alternatives, asymptotic theory is used as the basis for model development, but the issue of data scarcity leads to interesting challenges for creating models that optimise such data as are available.

Syllabus:

- Asymptotic theory for maxima of univariate independent and identically distributed (iid) random variables: limit distributions, GEV distribution, and domains of attraction
- Extension of asymptotic theory for univariate iid variables to cover top order statistics and threshold exceedances: GP distribution
- Statistical modelling and inference using maxima and threshold methods
- Statistical modelling of extremes of non-identically distributed random variables
- Asymptotic theory and statistical methods for extreme values of stationary sequences: clustering, extremal index

Bibliography:

S G Coles, An Introduction to the Statistical Modelling of Extreme Values, Springer-Verlag, London, 2001.

MATH563: Clinical Trials

Objectives of the module:

This course aims to introduce students to aspects of statistics, which are important in the design and analysis of clinical trials. On completion of the module students should understand the basic elements of clinical trials, be able to recognise and use principles of good study design, and be able to analyse and interpret study results to make correct scientific inferences.

Description

Clinical trials are planned experiments on human beings designed to assess the relative benefits of one or more forms of treatment. For instance, we might be interested in studying whether aspirin reduces the incidence of pregnancy-induced hypertension; or we may wish to assess whether a new immunosuppressive drug improves the survival rate of transplant recipients.

Syllabus:

This module combines the study of technical methodology with discussion of more general methodological and ethical issues:

- Clinical trials fundamentals: design issues, ethics and defining and estimating treatment
- Cross-over trials
- Sample size determination
- Equivalence and non-inferiority trials
- Meta-analysis

Bibliography:

D.G. Altman, Practical Statistics for Medical Research, Chapman and Hall, 1991.

S. Senn, Cross-over trials in clinical research, Wiley, 1993.

S. Piantadosi, Clinical Trials: A Methodologic Perspective, John Wiley & Sons, 1997.

ICH Harmonised Tripartite Guidelines.

J.N.S. Matthews, Introduction to Randomised Controlled Clinical Trials, Arnold, 2000.

MATH564: Principles of Epidemiology

Objectives of the module:

Epidemiology is the study of the distribution and determinants of disease in human populations. This course provides an introduction to the principles and statistical methods of epidemiology. Various concepts and strategies used in epidemiological studies are examined. Most inference will be likelihood based, although the emphasis is on conceptual considerations.

Syllabus:

- The history of epidemiology and the role of statistics
- Measures of health and disease: incidence, prevalence and cumulative incidence risk
- Types of epidemiological studies: Randomized controlled trials, cohort studies, case-control studies, cross-sectional and ecological studies
- Causation in epidemiology
- Potential errors in epidemiological studies: selection bias, confounding
- Remedies for confounding: Standardized rates, stratification and matching
- Diagnostic test studies

Bibliography:

R. Beaglehole, R. Bonita and T. Kjellstroem (1993) *Basic epidemiology*. Geneva: World Health Organization.

D. Clayton and M. Hills (1993) *Statistical models in epidemiology*. Oxford: Oxford University Press.

M. Woodward (1999) *Epidemiology: Study design and data analysis*. Chapman & Hall, Boca Raton

K.J. Rothman, S. Greenland and T.L. Lash. Modern Epidemiology. Lippincott Williams & Wilkins,US, 2008.

CHIC565: Environmental Epidemiology

Prerequisites: MATH551, MATH552, MATH564

Objectives of the module:

This course aims to introduce students to the kinds of statistical methods commonly used by epidemiologists and statisticians to investigate the relationship between risk of disease and environmental factors. Specifically the course concentrates on studies with a spatial component. A number of published studies will be used to illustrate the methods described, and students will learn how to perform similar analyses using the statistical R package. By the end of the course students should have an awareness of the kinds of methods used in environmental epidemiology, including an appreciation of their limitations. They should also be capable of conducting a number of these methods themselves.

Syllabus:

- Introduction: Motivating examples for methods in course.
- Spatial Point Processes: Theory and methods for the analysis of distributions of points in two-dimensional space; the Poisson process; univariate and bivariate K-functions.
- Spatial variation in risk: Case-control point-based methods and methods based on counts; spatial clustering.
- Disease mapping: Graphical investigation of spatial variation in risk; constructing smooth maps of disease risk from area-level count data.
- Geographical correlation studies: Poisson regression; the ecological fallacy and its relation with disease mapping.
- Point source methods: Investigation of risk associated with distance from a point or line source, for point and count data; practical implementation in epidemiological studies.

Bibliography:

P.J. Diggle. Statistical Analysis of Spatial Point Patterns (2nd edition). London: Edward Arnold. 2003.

P. Elliott, J. Cuzick, D. English and R. Stern (eds). Geographical and environmental epidemiology: methods for small-area studies. Oxford: Oxford University Press, 1992.

O. Schabenberger and C.A. Gotway. Statistical Methods in Spatial Data Analysis. Boca Raton: Chapman & Hall/CRC, 2005.

L. Waller and C.A. Gotway. Applied Spatial Statistics for Public Health Data. New York: Wiley, 2004.

MATH566: Longitudinal Data Analysis

Prerequisites: MATH551, MATH552

Objectives of the module:

Longitudinal data arise when a time-sequence of measurements is made on a response variable for each of a number of subjects in an experiment or observational study. For example, a patient's blood pressure may be measured daily following administration of one of several medical treatments for hypertension. The practical objective of many longitudinal studies is to find out how the average value of the response varies over time, and how this average response profile is affected by different experimental treatments. This module presents an approach to the analysis of longitudinal data, based on statistical modelling and likelihood methods of parameter estimation and hypothesis testing.

The specific aim of this module is to teach students a modern approach to the analysis of longitudinal data. Upon completion of this course the students should have acquired, from lectures and practical classes, the ability to build statistical models for longitudinal data, and to draw valid conclusions from their models.

Syllabus:

- What are longitudinal data?
- Exploratory and simple analysis strategies
- The independence working assumption
- Normal linear model with correlated errors
- Linear mixed effects models
- Generalised estimating equations
- Dealing with dropout

Bibliography:

H. Brown and R. Prescott, Applied Mixed Models in Medicine, Wiley, 1999.

P.J. Diggle, P. Heagerty, K.Y. Liang and S.L. Zeger, Analysis of Longitudinal Data (second edition), Oxford University Press, 2002.

G. Verbeke and G. Molenberghs, Linear Mixed Models for Longitudinal Data, Springer, 2000.

R. E. Weiss, Modelling longitudinal data, Springer, 2005.

MATH572: Genomics: technologies and data analyses

Objectives of the module:

To describe several modern genomics technologies in their biological context, to describe the types of data that are obtained from these technologies, to describe the statistical methodologies used to analyse such data and to have the students use statistical packages to perform such analyses on example data sets and interpret the results.

Description:

Genomics is a large field dealing with everything from DNA to metabolites and from evolution to microbiology. This course focuses on several genomics technologies by first of all putting them into their biological and genetic context and secondly describing the types of biological questions that can be answered with the data from these technologies. As an example of technologies that are to be discussed during this course are (i) DNA sequencing (ii) SNP (iii) microarrays and (iv) blotting and other proteomics methods. The most commonly use statistical analysis tools for each technology described will be discussed. The types of biological questions that are going to be addressed relate to issues such as genomic variation, constant genome and changing expression, human evolution and migration, disease and normality.

Syllabus:

- DNA sequencing, Single Nucleotide Polymorphisms (SNPs), transcriptional analysis via microarrays and Proteomics.
- Visualisation methods, hypothesis testing, multiple testing problems, multivariate methods, regression analyses for high dimensional data.

Bibliography:

Malcolm Campbell and Laurie J. Heyer. *Genomics, Proteomics and Bioinformatics*. CSHL Press, 2003.

Lange, K. Mathematical and Statistical Methods for Genetic Analysis, Springer, 2nd ed. 2002.

Wit, E. C. and McClure, J. D. Statistics for Microarrays: Design, Analysis and Inference, John Wiley & Sons, 2004.

MATH573: Survival and Event History Analysis

Prerequisites: MATH551, MATH552

Objectives of the module:

To describe the theory and to develop the practical skills required for the design and analysis of medical studies leading to the observation of survival times or multiple failure times. By the end of the course students should be able to develop study designs and to carry out sophisticated analyses of this type, should be aware of the variety of statistical models and methods now available, and understand the nature and importance of the underlying model assumptions.

Description:

In many medical applications interest lies in times to or between events. Examples include time from diagnosis of cancer to death, or times between epileptic seizures. This advanced course begins with a review of standard approaches to the analysis of possibly censored survival data. Survival models and estimation procedures are reviewed, and emphasis is placed on the underlying assumptions, how these might be evaluated through diagnostic methods and how robust the primary conclusions might be to their violation. Study design is considered, in particular how to define failure and censoring and how to determine a suitable sample size and duration of follow-up.

The course closes with a description of models and methods for the treatment of multivariate survival data, such as repeated failures, the lifetimes of family members or competing risks. Stratified models, marginal models and frailty models are discussed.

Syllabus:

- Survival data. Censoring. Survival, hazard and cumulative hazard functions. Kaplan-Meier plots. Parametric models and likelihood construction. Cox's proportional hazards model, partial likelihood
- Time-dependent covariates. Diagnostic methods. Residual analysis. Testing the proportional hazards assumption
- Competing risks data, cause-specific hazard and cumulative incidence functions
- Stratified models, marginal models, frailty models

Bibliography:

Collett, D. Modelling Survival Data in Medical Research. Chapman and Hall/CRC, 2003.

Hougaard, P. Analysis of Multivariate Survival Data. Springer, 2000.

Therneau, TM and Grambsch, PM. Modelling Survival Data: Extending the Cox Model. Springer, 2000.

Pintilie, M. Competing Risks: A Practical Perspective. Wiley, 2006.

Forecasting

Objectives

The module introduces time series and causal forecasting methods so that passing students will be able to prepare methodologically competent, understandable and concisely presented reports for clients. By the end of the course, students should be able to model causal and time series models, assess their accuracy and robustness and apply them in a real world problem domain.

Syllabus:

Introduction to Forecasting in Organisations:
- Extrapolative vs. Causal Forecasting
- History & academic research in Forecasting
- Forecasting case studies

Data Exploration:
- Time Series Patterns
- Univariate & Multivariate Visualisation
- Naïve Forecasting Methods & Averages

Measuring accuracy

Exponential Smoothing Methods:
- Single, Seasonal & Trended Exponential Smoothing
- Model Selection
- Parameter Selection

ARIMA Methods
- AR-, MA-, ARMA and ARIMA Models
- ARIMA Model specification & estimation
- Automatic selection

Time Series Regression
- Simple & multiple regression on time series
- Hypothesis testing,
- Model evaluation
- Diagnostics

Time Series Regression
- Model specification and constraints
- Dummy Variables, Lag, Non-linearities
- Stationarity
- Building regression models

Applications in operations and marketing

Judgmental Forecasting
- Judgmental methods for forecasting
- Biases and heuristics

Bibliography:

Ord K. & Fildes R. (2013), Principles of Business Forecasting, South-Western Cengage Learning.

Data Mining for Marketing, Sales and Finance

Objectives

The course extends the concepts of statistical model building and the models from the Introductory Statistics module towards methods from machine learning and artificial intelligence.

By the end of the course you should be able to:

- Understand general modelling concepts in relation to complex models
- Use a wide range of data mining methods to handle data of different types & applications
- Understand how these methods may be applied in practical management contexts
- Use & apply SAS Enterprise Miner to deal with complexity and large datasets

Syllabus

- Introduction to Data Mining
- Data Mining Process
    - Methods for data exploration & manipulation
    - Methods for data reduction & feature selection
    - Evaluating Classification Accuracy
- Data Mining Methods for Classification
    - Logistic Regression
    - Decision Trees
    - Nearest neighbour classification
    - Artificial Neural Networks
- Data Mining applications in Credit Scoring

Bibliography

- Tan, P. N., M. Steinbach, et al. (2005). Introduction to data mining. Boston, Pearson Addison Wesley.
- Berry, M. J. A. and G. Linoff (2000). Mastering data mining: the art and science of customer relationship management. New York, NY [u.a.], Wiley Computer Publ.
- Berry, M. J. A. and G. Linoff (2004). Data mining techniques: for marketing, sales, and customer relationship management. Indianapolis, Ind., Wiley Pub.
- Linoff, G. and M. J. A. Berry (2001). Mining the Web: transforming customer data into customer value. New York, John Wiley & Sons.
- Weiss, S. M. and N. Indurkhya (1998). Predictive data mining: a practical guide. San Francisco, Morgan Kaufmann Publishers.

Optimisation and Heuristics

Objectives

By the end of the course you should be able to:

- know how to formulate problems as mathematical programs and solve them;
- be aware of the power, and the limitations, of optimisation methods;
- be able to carry out sensitivity analysis to see how robust the recommendation is;
- be familiar with commercial software such as MPL, LINDO and EXCEL SOLVER;
- be aware of major heuristic techniques and know when and how to apply them.

Introduction

Optimisation, sometimes called *mathematical programming*, has applications in many fields, including Operational Research, Computer Science, Statistics, Finance, Engineering and the Physical Sciences. Commercial optimisation software is now capable of solving many industrial-scale problems to proven optimality. On the other hand, there are still many practical applications where finding a provably-optimal solution is not computationally viable. In such cases, *heuristic* methods can allow good solutions to be found within a reasonable computation time.

The course is designed to enable students to apply optimisation techniques to business problems. Building on the introduction to optimisation in MSCI502 and/or MSCI519, students will be introduced to different problem formulations and algorithmic methods to guide decision making in business and other organisations.

Syllabus

- Linear Programming
- Non-Linear Programming
- Integer and Mixed-Integer Programming
- Dynamic Programming
- Heuristics

Bibliography

HP Williams (2013) *Model Building in Mathematical Programming* (5th edition). Wiley. ISBN: 978-1-118-44333-0 (pbk).

J Kallrath & JM Wilson (1997) *Business Optimisation Using Mathematical Programming*. Macmillan. ISBN: 0-333-67623-8.

WL Winston (2004) *Operations Research - Applications and Algorithms* (4th edition). Thompson. ISBN: 978-0534380588.

DR Anderson, DJ Sweeney, TA Williams & M. Wisniewski (2008) *An Introduction to Management Science*. Cengage Learning. ISBN: 978-1844805952.

E.K. Burke & G. Kendall (eds.) (2005) *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*. Springer.