# MSc Statistics (and Streams) Module Guide 2025-26

## Module schedule

| Term | Week | Module Code | Module Name | Type | ECTS |
|---|---|---|---|---|---|
| Autumn | 2-11 | MATH70071 | Applied Statistics | Core | 7.5 |
| Autumn | 2-11 | MATH70078 | Fundamentals of Statistical Inference | Core | 7.5 |
| Autumn | 2-11 | MATH70082 | Probability for Statistics | Core | 7.5 |
| Autumn | 2-11 | MATH70093 | Computational Statistics and Machine Learning | Core | 7.5 |
| Autumn & Spring | All | MATH70076 | Data Science | Core | 5 |
| Spring | 2-6 | MATH70079 | Introduction to Statistical Finance | Elective | 5 |
| Spring | 2-6 | MATH70089 | Stochastic Processes | Elective | 5 |
| Spring | 2-6 | MATH70013 | Advanced Simulation Methods | Elective | 5 |
| Spring | 7-11 | MATH70081 | Nonparametric Statistics | Elective | 5 |
| Spring | 7-11 | MATH70070 | Advanced Statistical Finance | Elective | 5 |
| Spring | 7-11 | MATH70083 | Statistical Learning for High-Dimensional Data | Elective | 5 |
| Spring | 1-11 | MATH70046 | Time Series Analysis | Elective | 7.5 |
| Spring | 1-11 | MATH70048 | Survival Models | Elective | 7.5 |
| Spring | 1-11 | MATH70099 | Big Data: Statistical Scalability with PySpark | Elective | 5 |
| Spring | 1-11 | MATH70101 | Deep Learning | Elective | 7.5 |
| Spring | 1-11 | MATH70139 | Spatial Statistics | Elective | 7.5 |
| Spring | 1-11 | MATH70148 | Probabilistic Generative Models | Elective | 7.5 |

The information displayed in this guide is correct at the time of publication and is subject to change.

# Module Descriptions

## Core Modules

### MATH70071: Applied Statistics

**Module Leader:** Prof D. Mortlock

**Description**

The module focuses on statistical modelling and regression when applied to realistic problems and real data. This module will explore the following classes of statistical models: Linear Models, Generalized Linear Models, Normal Linear Mixed Models. For each class of models we will look at: mathematically defining the model, estimation procedures, inference and diagnostics from both frequentist and Bayesian approaches.

**Learning Outcome**

On successful completion of this module students will be able to:

- Deploy statistical software packages to perform data analyses;
- Understand the mathematical foundations and practical challenges of statistical modelling, including but not limited to normal linear models and generalized linear models;
- Finesse statistical models to appropriately and directly address substantive questions arising in data analyses;
- Understand how to interpret models fitted to data in terms of the substantive setting;
- Formulate statistical models in order to leverage their mathematical foundations to address real applied data-analytic problems;
- Gain intuition for statistical data analysis and modelling from practical experience.

**Module Content**

This module will explore a variety of signature statistical modelling techniques that are widely used across all major branches of applied statistics, including but not limited to Linear Models, Generalized Linear Models. For each class of model we will look at: mathematically defining the model, estimation procedures, inference and diagnostics.

**Assessment:** 40% in-class tests,60% coursework.

### MATH70076: Data Science

**Module Leaders:** Dr Z. Varty and Dr. R. Nuermaimaiti

**Description**

Model building and evaluation are necessary but not sufficient skills for the effective practice of data science. In this module students will develop the broader set of skills required of a successful data scientist within an organisation. During the module students will explore and critique how to: effectively scope and manage a data science project; work openly and reproducibly; efficiently acquire, manipulate, and present data; interpret and explain models to a variety of stakeholders; assess a model's readiness for production; and appraise the ethical implications of their work as a data scientist. The module has a practical focus, with an emphasis on learning through student activities and group discussion.

**Learning Outcome**

On successful completion of this module students will be able to:

- Plan data analysis and data gathering techniques;
- Produce and write reproducible workflows;
- Plan and explain preparation and exploration of messy real-world datasets;
- Create statistical graphics to visualize and present data;
- Write about and explain their data analysis procedure; and
- Explain, critique, judge and think critically about data scientific approaches.

**Module Content**

The year-long structure of the module recognizes that students will in the context of the overall MSc benefit from achieving learning outcomes at specific checkpoints. Specifically, at the start of term 1 students will benefit from taught components on reproducible data science workflows and academic integrity, before the other four core modules begin. At the start of term 2, students will benefit from taught components on data acquisition and data exploration. They will also learn about data science ethics, and critical thinking about appropriate uses and misuses of statistical science. Over term 3 and the summer, students will benefit from hands-on training on debugging and benchmarking, finishing up research for dissemination, and project presentation before they complete their research project.

**Assessment:** 100% coursework.

## MATH70078: Fundamentals of Statistical Inference

**Module Leader:** Prof A. Young

**Description**

The module provides a rigorous introduction to key aspects of statistical inference, describing Bayesian, frequentist and Fisherian approaches to point estimation, hypothesis testing and confidence set construction. Its focus is on comparison of the different approaches in parametric problems, in both small and large sample settings.

**Learning Outcome**

On successful completion of this module students will be able to:

- Describe and appraise the key aspects of different Bayesian and frequentist approaches to statistical inference;
- Evaluate and justify appropriate methods for hypothesis testing, point estimation and confidence set construction in parametric settings, for both small sample and large sample data problems;
- Apply statistical theory to explain the properties of inferential procedures;
- Critically evaluate the operational characteristics of competing inferential methods;
- Explain the key principles underlying statistical theory and methodology."

**Module Content**

In statistical inference experimental or observational data are modelled as the observed values of random variables, to provide a framework from which inductive conclusions may be drawn about the mechanism giving rise to the data. This is done by supposing that the random variable has an assumed parametric probability distribution: the inference is performed by assessing some aspect of the parameter of the distribution.

This module develops the main approaches to statistical inference for point estimation, hypothesis testing and confidence set construction. Focus is on description of the key elements of Bayesian, frequentist and Fisherian inference through development of the central underlying principles of statistical theory. Formal treatment is given of a decision-theoretic formulation of statistical inference. Key elements of Bayesian and frequentist theory are described, focussing on inferential methods deriving from important special classes of parametric problem and application of principles of data reduction. General purpose methods of inference deriving from the principle of maximum likelihood are detailed. Throughout, particular attention is given to evaluation of the comparative properties of competing methods of inference.

**Assessment:** 75% written exam, 25% coursework.

## MATH70082: Probability for Statistics

**Module Leader:** Dr Y. Tang

### Description

This module introduces the key concepts of probability theory in a rigorous fashion, with the aim of providing a foundational core for the optional modules, starting from a rigorous definition of probability spaces, random variables and probability distributions. Concepts such as independence and convergence of random variables are presented, leading to limit theorems, including laws of large numbers and central limit theorems. The second part of the module will focus on sequences of dependent random variables, leading up to discrete time Markov processes and the long-time behaviour of such processes.

### Learning Outcome

On successful completion of this module students will be able to:

- Understand the key elements of rigorous probability theory.
- Analyse random models, defined in terms of constituent random variables and distributions,
- Analyse the dependence between random models through the various notions of independence.
- Understand the theory underpinning Lebesgue-Stieljes integration, and apply it to provide a rigorous meaning to expectations and conditional expectations of random variables.
- Understand the main modes of convergence of random variables, and remember relevant examples and counterexamples which distinguish between theorems
- Apply the law of large numbers and central limit theorem to analyse the asymptotic properties of various random models arising in statistics.
- Analyse the properties of various classes of dependent sequences of random variables.
- Understand key properties of discrete time Markov chains, including the CK-equations, classification of states, recurrence and transciences, stationarity, time reversibility and ergodicity.
- Analysing the long term behaviour of a Markov chain, and evaluating the properties of its stationary distributions (if they exist).

### Module Content

The module Probability for Statistics introduces the key concepts of probability theory in a rigorous way. Topics covered include: the elements of a probability space, random variables and vectors, distribution functions, independence of random variable/vectors, a concise review of the Lebesgue-Stieltjes integration theory, expectation, modes of convergence of random variables, law of large numbers, central limit theorems, characteristic functions, conditional probability and expectation.

**Assessment:** 75% written exam, 25% coursework

## MATH70093: Computational Statistics and Machine Learning

**Module Leader:** Dr A. Duncan and Dr S. Filippi

**Description**

This module covers a number of computational methods that are key in modern statistics. The module will focus on statistical computing including deterministic and probabilistic numerical methods from root finding, numerical integration, optimisation to Monte Carlo integration, bootstrapping, and Markov Chain Monte Carlo, as well as an introduction to machine learning that will cover basic learning theory and canonical neural network architectures.

**Learning Outcome**

On successful completion of this module students will be able to:

- Explain several fundamental methods for statistical computing;
- Implement computational methods confidently;
- Apply and adapt state-of-the-art methods in computational statistics to a specific dataset;
- Analyse the relative advantages and limitations of computational and numerical methods.

**Module Content**

This module covers a number of computational methods that are key in modern statistics. The module will focus on statistical computing including deterministic and probabiilistic numerical methods from root finding, numerical integration, optimisation to Monte Carlo integration, bootstrapping, and Markov Chain Monte Carlo, as well as an introduction to machine learning that will cover basic learning theory and canonical neural network architectures.

**Assessment:** 40% in-class tests, 60% coursework.

## Elective - 5 ECTS Modules

### MATH70013: Advanced Simulation Methods

**Module Leader:** Dr N. Kantas

**Description**

Modern problems in statistics require sampling from complicated probability distributions defined on a variety of spaces and setups. In this module we will visit popular advanced sampling techniques, such as Importance Sampling, Markov Chain Monte Carlo, Sequential Monte Carlo. We will consider the underlying principles of each method as well as practical aspects related to implementation, computational cost, and efficiency. By the end of the module the students will be familiar with these sampling methods and will have applied them to popular models, such as Hidden Markov Models, which appear ubiquitous in many scientific disciplines.

**Learning Outcome**

On successful completion of this module students will be able to:

- Create, evaluate, understand and analyse advanced simulation methods;
- Display mastery of Hidden Markov models and what simulation methods are required to fit them to data;
- Understand, apply and evaluate variance and bias assessment and reduction;
- Display mastery of advanced Monte Carlo methods used in Statistics, such as Importance Sampling, Markov Chain Monte Carlo, Sequential Monte Carlo;
- Understand the basics and foundations of the methods mentioned above. Effort will be made to link the principles behind Monte Carlo methods and practical problems.

**Module Content**

This module will explore simulation algorithms: Importance Sampling, Markov Chain Monte Carlo, Sequential Monte Carlo and Particle filtering. We will discuss Hidden Markov Models, the simulation of the conditional laws and static parameter estimation. Popular examples: Linear Gaussian models, Stochastic Volatility, SIR models in epidemics etc.

**Assessment:** 100% coursework.

## MATH70070: Advanced Statistical Finance

**Module Leader:** Dr M. Pakkanen

### Description

Advanced Statistical Finance focuses on modern statistical methods for analysis of financial data. During the last two decades, the increasing availability of large financial data sets has prompted development of new statistical and econometric methods that can cope with high-dimensional data, high-frequency observations and extreme values in data. The module presents the elements of extreme value theory, continuous-time financial modelling, volatility estimation and forecasting, and high-dimensional covariance matrix estimation.

### Learning Outcome

On successful completion of this module students will be able to:

- Explain the asymptotic behaviour of maxima of random variables and apply this theory to model extremes in financial data;
- Design continuous-time stochastic models for high-frequency financial data using stochastic integration and Itô calculus;
- Apply realised measures, such as the realised variance, to analyse and draw inference on the empirical properties of realised volatility in high-frequency financial data;
- Derive volatility forecasts using models that incorporate realised measures;
- Analyse the accuracy of volatility forecasts using modern forecast evaluation methods;
- Develop accurate estimators of the covariance matrix of asset returns when the number of assets exceeds the time-dimension of the data set.

### Module Content

The module will first introduce the basics of extreme value theory, which will be used to develop models and estimation methods for extremes in financial data.

The second part of the module will provide a concise introduction to the theory of stochastic integration and Itô calculus, which provide a theoretical foundation for volatility estimation from high-frequency data using the concept of realised variance. The asymptotic properties of realised variance will be elucidated and applied to draw inference on realised volatility.

The third part introduces some recently developed volatility forecasting models that incorporate volatility information from high-frequency data and demonstrates how the performance of such models can be assessed and compared using modern forecast evaluation methods such as the Diebold-Mariano test and the model confidence set.

The final part of the module provides an overview of covariance matrix estimation in a high-dimensional setting, motivated by applications to variance-optimal portfolios. The pitfalls of using the standard sample covariance matrix with high-dimensional data are first exemplified. Then it is shown how shrinkage methods can be applied to estimate covariance matrices accurately using high-dimensional data.

**Assessment:** 100% written exam.

## MATH70079: Introduction to Statistical Finance

**Module Leader:** Dr A. Luati

**Description**

Introduction to Statistical Finance presents the fundamental concepts of quantitative finance and statistical methods that are widely used to analyse financial data. It starts off with a concise introduction to financial markets, proceeding then to the basic principles of derivatives pricing and risk measurement. Subsequently, it develops statistical models for financial time series. Finally, it explains how such models are estimated, how their goodness of fit can be assessed and how they can be used in forecasting.

**Learning Outcome**

On successful completion of this module students will be able to:

- explain the risk-neutral pricing theory widely used in quantitative finance;
- explain the concept of risk measures;
- define ARMA-GARCH processes and analyse their key properties;
- design estimation procedures for ARMA-GARCH processes;
- derive formulas for value at risk and expected shortfall for ARMA-GARCH processes;
- apply ARMA-GARCH processes to empirical data and analyse their goodness-of-fit;
- derive forecasts for financial time series;

**Module Content**

The module introduces fundamental concepts in financial economics and quantitative finance and presents suitable statistical tools which are widely used when analysing financial data. The module will start off with an introduction to risk-neutral pricing theory followed by a primer on risk measures such as value at risk and expected shortfall which are widely used in financial risk management. Next, an introduction to time series analysis will be given, where the focus will be on so-called ARMA-GARCH processes. Such processes can describe some of the stylised facts widely observed in financial data, including non-Gaussian returns and heteroskedasticity. Finally, methods for forecasting financial time series will be introduced.

**Assessment:** 100% written exam.

## MATH70081: Nonparametric Statistics

**Module Leader:** Dr K. Ray

**Description**

Nonparametric methods aim to provide inference under weaker assumptions than conventional parametric methods. In this module, students will apply modern techniques to a variety of problems, such as estimating distribution functions, density estimation and nonparametric regression.

**Learning Outcome**

On successful completion of this module students will be able to:

- Evaluate the strengths and weaknesses of nonparametric methods as flexible data-driven alternatives to parametric modelling;
- Describe basic properties of the empirical distribution function;
- Describe the bias-variance trade-off for nonparametric estimation;
- Construct nonparametric density estimators, for example kernel density estimators, and appraise their relative merits using common statistical criteria, such as mean square error, bias and variance;

- Construct nonparametric regression estimators, for example Nadaraya-Watson estimators or cubic splines, and appraise their relative merits using common statistical criteria;
- Appreciate basis functions, for example B-splines or wavelets, as an efficient method of representing functions and understand how they can be used to construct nonparametric estimators.

**Module Content**

The module will first introduce the most basic nonparametric estimator, the empirical distribution function, and cover its basic properties and how it can be used to construct confidence intervals. We then consider the problem of nonparametric density estimation, illustrating the key concept of the bias-variance trade-off via histograms and kernel density estimators. To evaluate the statistical properties of the latter, we will examine approximation by convolutions.

The module then covers various methods in nonparametric regression, such as kernel methods, local polynomial regression or splines. We will finally turn to the basis function approach, using suitable bases to provide efficient function approximations and hence derive estimators. This will be illustrated via B-splines and wavelets.

**Assessment:** 100% written exam.

## MATH70083: Statistical Learning for High-Dimensional Data

**Module Leader:** Dr M. Evangelou

**Description**

This module introduces different models and tools used for the analysis of complex and high-dimensional datasets, like the ones observed in the field of genetics.

**Learning Outcome**

On successful completion of this module students will be able to:

- Appreciate the importance of handling high dimensional datasets, through the analysis of high dimensional datasets from different fields including genetics
- Define multiple testing approaches and prove related theorems;
- Derive statistical approaches, including Bayesian and Frequentist, for performing variable selection for high dimensional data
- Evaluate different model selection approaches

**Module Content**

In this module we will develop models and tools to analyse complex and high dimensional datasets as these arise in different application fields as for example the genetics field. This will include statistical and machine learning techniques for multiple testing, penalised regression, clustering, dimensionality reduction and visualisation. The module will cover both Frequentist and Bayesian statistical approaches as well as some modern machine learning unsupervised and supervised approaches.

**Assessment:** 30% in-class test, 70% coursework

## MATH70089: Stochastic Processes

**Module Leader:** Dr Randolf Altmeyer

**Description**

A stochastic process is a statistical model for describing phenomena that evolve dynamically in a random manner over time. This module will cover aspects of continuous-time stochastic processes, with a focus on theory, and with some computation. Various classes of stochastic processes will be considered, but the Wiener process and diffusions will be emphasised. Multiple applications to statistical finance will also be presented.

**Learning Outcome**

On successful completion of this module, students will be able to:

- Understand the key principles underlying the theory of stochastic processes in continuous time, in particular the theory of the Wiener process.
- Analyse properties of stochastic process models, identifying notions such as markovianity, the martingale property, and stopping times.
- Understand the construction of key classes of stochastic processes, including Brownian motion and diffusions. Ito's formula and its key related results should also be understood.
- Create code for simulating continuous-time stochastic processes.
- Understand applications of diffusions in statistical finance.

**Module Content**

The goal of the module is to provide a masters-level introduction to continuous-time stochastic processes, with a focus on conceptual understanding. The module will introduce martingales, continuous-time Markov processes, stopping times, the Wiener process, Ito diffusion and more general Gaussian processes. The module will also provide a gentle introduction to stochastic differential equations, starting from stochastic integrals and leading onto a general definition of diffusion processes through Ito's formula and related results. Key applications in statistical finance will also be covered.

**Assessment:** 100% written exam. [KT1][MR2]

## MATH70099: Big Data: Statistical Scalability with PySpark

**Module Leader:** Dr F. Sanna Passino

**Description**

The module consists of three components: statistical analysis at scale, distributed programming using MapReduce, Big Data analysis using PySpark. The first component covers theory on statistical scalability, and discusses topics such as sufficiency, Statistical Query Model, stochastic and distributed optimisation, stochastic variational inference, Markov Chain Monte Carlo methods for tall data, and statistical analysis of streaming data. The second and third components cover practical aspects of handling Big Data, introducing two frameworks for analysis of large datasets: Hadoop and Spark.

**Learning Outcome**

- Extract Transform and Load data using using Hadoop Distributed File System in order to load data into and out of a big data environment.
- Use PySpark interface in order to interact with huge data sets.
- Perform EDA using PySpark in order to understand underpinning statistical properties of data set being analysed.
- Explain how underpinning statistical methodology can be applied to big data.
- Apply underpinning statistical methodology to big data using Pyspark, in order to be able to produce statistical conclusion.
- Combine EDA methodology and PySpark knowledge to produce full statistical analysis on huge data sets.

**Module Content**

1. Introduction to Big Data (history and characteristics) and statistical methods for Big Data analysis (divide and conquer, subsampling)
2. Introduction to the Hadoop Distributed File System (HDFS) and MapReduce-based processing
3. Statistical modelling in MapReduce (sufficiency and Statistical Query Model)
4. Introduction to Resilient Distributed Datasets (RDDs) and DataFrames in PySpark for large scale Big Data Processing
5. Statistical modelling with PySpark in the library ML
6. User-defined functions and additional libraries in PySpark
7. Optimisation with Big Data (stochastic gradient algorithms and stochastic variational inference)
8. Markov Chain Monte Carlo and Big Data (divide-and-conquer MCMC and subsampling-based methods)
9. Streaming data analysis (forgetting factor methods, exponentially weighted moving averages, change detection methods)

**Assessment:** 100% coursework.

## Elective - 7.5 ECTS Modules

### MATH70046 Time Series Analysis

**Module Leader:** Dr E. Cohen

**Description**

A time series is a series of data points indexed and evolving in time. Importantly, we cannot assume the observations are independent from one another and, in fact, they are often significantly correlated. They are prevalent in many areas of modern life, including science, engineering, business, economics, and finance. This module is a self-contained introduction to the analysis of time series. Weight is given to both the time domain and frequency domain viewpoints, and important structural features (e.g., stationarity, invertibility) are treated rigorously. Attention is given to modelling, estimation and prediction (forecasting), and useful computational algorithms and approaches are introduced.

**Learning Outcome**

On successful completion of this module, you will be able to:

- Appreciate that time series should be considered observations from an underlying stochastic process.
- Define what it means for a time series to be stationary.
- Identify autocorrelation within time series data.
- Work with standard models of time series.
- Appreciate that time series can exhibit trend and seasonality and know how to adjust for these.
- Determine the spectral representation of stationary time series and use the spectral density function to provide an alternative viewpoint of second-order structure.
- Derive and implement estimators of mean, correlation and spectral properties.
- Extend time series models, the notion of stationarity, and frequency domain representations to multivariate time series.
- Derive forecasts from standard time series models and quantify their uncertainty.
- Demonstrate an integrated understanding of the concepts of this module by independent study of related material.

**Module Content**

An indicative list of sections and topics is:

- Discrete time stochastic processes and examples.
- Autocovariance, autocorrelation, stationarity.
- Trend removal and seasonal adjustment.
- AR, MA and ARMA processes, characteristic polynomials, general linear process, invertibility, directionality and reversibility.
- Spectral representation, aliasing, linear filtering.
- Estimation of mean and autocovariance sequence, the periodogram, tapering for bias reduction.
- Parametric model fitting.
- Forecasting.
- Independent study of extension material (in the form of a book chapter, additional notes or a research paper) applying or extending material from the above topics.

**Assessment:** 100% written exam

## MATH70048 Survival Models

**Module Leader:** Dr R. Nuermaimaiti

**Description**

Survival models are fundamental to actuarial work, as well as being a key concept in medical statistics.

This module will introduce the ideas, placing particular emphasis on actuarial applications.

**Learning Outcome**

On successful completion of this module, you will be able to:

- Appreciate survival analysis models as a framework that deals with lifetimes and censored observations;
- Use several methods to define event time distributions and the relation between these definitions;
- Describe, select and use methods for fitting parametric, semi-parametric and non-parametric survival analysis models, including regression models and multi state models.
- Explain the counting process approach to survival analysis and its benefits;
- Use and critically analyse methods occurring in actuarial applications, such as methods for the construction and use of life tables.
- Demonstrate an integrated understanding of the concepts of this module by independent study of related material.

**Module Content**

An indicative list of sections and topics is:

1. Concepts of survival models,
2. Right and left censored data and randomly censored data.
3. Estimation procedures for lifetime distributions:
4. Empirical survival functions,
5. Kaplan-Meier estimates,
6. Cox model.
7. Statistical models of transfers between multiple states,
8. Maximum likelihood estimators.

9. Counting process models.
10. Actuarial Applications:
11. Life table data and expectation of life,
12. Binomial model of mortality,
13. The Poisson model,
14. Estimation of transition intensities that depend on age,
15. Graduation and testing of crude and smoothed estimates for consistency.
16. Independent study of extension material (in the form of a book chapter, additional notes or a research paper) applying or extending material from the above topics.

**Assessment:** 90% written exam, 10% coursework.

## MATH70101: Deep Learning

**Module Leader:** Dr K. Webster

### Description

This module teaches the building blocks of deep learning models, and how to design network architectures for specific applications, in both supervised and unsupervised contexts. It also covers practical skills in implementing neural networks. Students will learn how to design, implement, train and evaluate networks. A central focus of the module is on the mathematical and statistical foundations of some of the most sophisticated deep learning models, such as variational autoencoders (VAEs) and Bayesian methods for neural networks.

### Learning Outcome

- Select appropriate deep learning model architectures for given supervised and unsupervised learning applications.
- Implement different neural network model architectures, loss functions and optimisers using either the PyTorch or TensorFlow 2.x framework.
- Implement data and training pipelines for different types of neural networks using either the Tensorflow or PyTorch framework
- Implement appropriate evaluation measures and model selection strategies for supervised and unsupervised applications

### Module Content

1. Deep learning fundamentals, layers, activation functions, loss functions
2. Optimising deep learning models.
3. Backpropagation algorithm
4. Convolutional neural networks
5. Sequence models.
6. Recurrent neural networks
7. VAEs, generative models
8. Bayesian methods for deep learning.

**Assessment:** 100% coursework.

## MATH70139: Spatial Statistics

**Module Leader:** Dr A. Sykulski

### Description

Data collected in space are common in many applications including climate science, epidemiology, and economics. This module covers theoretical and methodological statistical foundations for spatial data. The module is structured to cover in detail the three fundamental forms in which spatial data are collected: gridded data, network data, and point pattern data.

For each data type, stochastic models will be defined and explored including random fields and point processes. Properties including isotropy, stationarity and homogeneity will be formalised and explored in the context of each model and data type. In addition, techniques for spatial interpolation will be studied.

**Learning Outcome**

On successful completion of this module students will be able to:

- Recognise the key differences between different types of spatial data (gridded, network and point pattern)
- Apply different classes of spatial covariance models to gridded and network data - Formulate different intensity functions for point pattern data
- Define the concepts of homogeneity, stationarity and isotropy in the context of spatial models - Select, derive and apply appropriate methods for spatial interpolation
- Demonstrate an integrated understanding of the concepts of the module by critical, independent study of research articles and books.

**Module Content**

1. Introduction to spatial data (gridded, network and point pattern data)
2. Random Fields and Covariance Functions
3. Spatial Interpolation and Kriging
4. Network Data and Markov Random Fields
5. Spatial Point Processes

**Assessment:** 90% written exam, 10% coursework.

## MATH70148 Probabilistic Generative Models

**Module Leader:** Dr F. Tobar

**Description**

Probabilistic generative models (PGM) are at the forefront of statistical machine learning research and are central in contemporary AI applications. This module develops a foundation for the design, analysis and implementation of PGMs. The module starts from the fundamentals of training and inference and then studies different PGMs, including parametric, autoregressive, explicit, implicit and (Bayesian) nonparametric models. Dissimilarity metrics are also examined, all to establish a comprehensive setting for the analysis of the models under study. Emphasising theory and practice, the module's assessment features two courseworks (20% each) and an exam (60%). Students are expected to be familiar with classic machine learning models and use Python in their assessments.

**Learning Outcomes**

On the successful completion of the module, you will be able to:

- select which generative model is appropriate for a given data analysis setup;
- train different types of generative models using dedicated Python toolboxes;
- calculate the similarity between different models;
- obtain samples from a trained model in an efficient and accurate manner;
- determine whether a training procedure has concluded successfully
- demonstrate an integrated understanding of the concepts of this module by independent study of related material.

**Module Content**

This module will cover the following topics:

1. Fundamentals: graphical models, exact and approximate Bayesian inference;
2. Dissimilarity measures: optimal transport, maximum mean discrepancy, notions of information theory;
3. Explicit models: normalising flows;
4. Variational autoencoders & generative adversarial networks;
5. Autoregressive models: linear filters, recurrent neural networks, transformers;
6. Bayesian nonparametric models: Dirichlet process & Gaussian process;
7. Diffusion models & score-based models

Assessment: 60% written exam, 40% coursework

## Streams

| General Stream | |
|---|---|
| Choose 25-27.5 ECTS | |
| 5 ECTS | 7.5 ECTS |
| 1. Advanced Simulation Methods<br>2. Advanced Statistical Finance<br>3. Big Data: Statistical Scalability with PySpark<br>4. Introduction to Statistical Finance<br>5. Nonparametric Statistics<br>6. Statistical Learning for High-Dimensional Data<br>7. Stochastic Processes | 8. Deep Learning<br>9. Survival Models<br>10. Time Series Analysis<br>11. Spatial Statistics<br>12. Probabilistic Generative Models |

| Applied Statistics Stream |
|---|
| Choose at least three modules from the modules listed below. Choose any other elective modules from the General Stream for a total of 25 - 27.5 ECTS worth of elective modules. |
| 1. Advanced Simulation Methods (5 ECTS)<br>2. Introduction to Statistical Finance (5 ECTS)<br>3. Statistical Learning for High-Dimensional Data (5 ECTS)<br>4. Probabilistic Generative Models (7.5 ECTS) |

| Biostatistics Stream |
|---|
| Choose at least three modules from the modules listed below. Choose any other elective modules from the General Stream for a total of 25 – 27.5 ECTS worth of elective modules. |
| 1. Advanced Simulation Methods (5 ECTS)<br>2. Big Data: Statistical Scalability with PySpark (5 ECTS)<br>3. Statistical Learning for High-Dimensional Data (5 ECTS)<br>4. Time Series (7.5 ECTS) |

| Data Science and Machine Learning Stream |
| --- |
| Choose at least three modules from the modules listed below. Choose any other elective modules from the General Stream for a total of 25 – 27.5 ECTS worth of elective modules. |
| 1. Deep Learning (7.5 ECTS)<br>2. Nonparametric Statistics<br>3. Big Data: Statistical Scalability with PySpark (5 ECTS)<br>4. Probabilistic Generative Models (7.5 ECTS) |

| Statistical Finance Stream |
| --- |
| Choose at least three modules from the modules listed below. Choose any other elective modules from the General stream for a total of 25 – 27.5 ECTS worth of elective modules. |
| 1. Advanced Simulation Methods (5 ECTS)<br>2. Big Data: Statistical Scalability with PySpark (5 ECTS)<br>3. Introduction to Statistical Finance (5 ECTS)<br>4. Advanced Statistical Finance (5 ECTS)<br>5. Stochastic Processes (5 ECTS)<br>6. Survival Models (7.5 ECTS) |

| Theory and Methods Stream |
| --- |
| Choose at least three modules from the modules listed below. Choose any other elective modules from the General stream for a total of 25 – 27.5 ECTS worth of elective modules. |
| 1. Advanced Simulation Methods (5 ECTS)<br>2. Stochastic Processes (5 ECTS)<br>3. Survival Models (7.5 ECTS)<br>4. Probabilistic Generative Models (7.5 ECTS)<br>5. Nonparametric Statistics (5 ECTS) |