MA3K0	Term 2
High-Dimensional Probability	15 CATS

Commitment: 30 hours

Prerequisites: ST111 Probability A & B; MA258 Mathematical Analysis III or MA259 Multivariate Calculus or ST208 Mathematical Methods or MA244 Analysis III; MA359 Measure Theory or ST342 Maths of Random Events. Earlier probability modules will be of some use. The framework is some mild probability theory (e.g., the following modules can be useful: ST342 Maths of Random Events, ST202 Stochastic Processes, MA3H2 Markov Processes and Percolation Theory). There are also strong links and thus suitable combinations to the following modules (MA4K4 Topics in Interacting Particle Systems, or MA4F7 Brownian Motions, or MA427 Ergodic Theory, or MA424 Dynamical Systems, or MA4L2 Statistical Mechanics, or MA4L2 Large deviation theory).

Content:

High-Dimensional Probability offers insight into the behaviour of random vectors, random matrices, random subspaces, and objects used to quantify uncertainty in high dimensions. Drawing on ideas from probability, analysis, and geometry, it lends itself to applications in mathematics, data science, machine learning, theoretical computer science, signal processing, optimisation, and more. Methods of high-dimensional probability have become indispensable in numerous problems of probability theory and its application in mathematics, data science & machine learning, and computer science.

The aim of the lecture is to integrate theory, key tools, and modern applications of high-dimensional probability. **Concentration inequalities** form the core of the module, and the lecture covers both classical results such as Hoeffding's and Chernoff's inequalities and modern developments such as the matrix Bernstein inequality. It then introduces powerful methods based on stochastic processes, including such tools as Slepian's, Sudakov's, and Dudley's inequalities, as well as generic chaining and various bounds. A broad range of illustrations will be discussed throughout, including classical and modern results for covariance estimation, clustering, networks, semidefinite programming, coding, dimension reduction, matrix completion, machine learning, and compressed sensing. A second core aim of the module is to study graphical models using methods from statistical mechanics (Ising model and large deviation and information theory) and deep learning (Boltzmann machines).

Aims:

1. Preliminaries on Random Variables (limit theorems, classical inequalities, Gaussian models, Monte Carlo)
2. Basic Information theory (entropy; Kull-Back Leibler information divergence)
3. Concentrations of Sums of Independent Random Variables
4. Random Vectors in High Dimensions
5. Random Matrices
6. Concentration with Dependency structures
7. Deviations of Random Matrices and Geometric Consequences
8. Graphical models and deep learning

Objectives: On successful completion of this module, a student should

1. Be able to understand the concentration of measure problem in high dimensions
2. Be able to distinguish three basic concentration inequalities
3. Be familiar with the variational principles
4. Familiar with random matrices (main properties)
5. Be able to distinguish between concentration for independent families as well as for various dependency structures
6. Understanding of basic concentrations of the norm
7. Be familiar with some application of graphical models

Leads to: ST343 Topics in Data Science, MA4L2 Large Deviation Theory, MA4F7 Brownian Motion, MA482 Stochastic Analysis, MA4G6 Calculus of Variations, MA4J5 Structures of Complex Systems

Books: We won't follow a particular book and will provide lecture notes. The course is based on the following three books where the majority is taken from [1]:

[1] Roman Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Cambridge Series in Statistical and Probabilistic Mathematics, (2018).

[2] Kevin P. Murphy, *Machine Learning - A Probabilistic Perspective*, MIT Press (2012).
[3] Simon Rogers and Mark Girolami, *A first course in Machine Learning*, CRC Press (2017).
[4] Alex Kulesza and Ben Taskar, *Determinantal point processes for machine learning* Lecture Notes (2013).

Assessment: 3-hour examination (85%). 15% homework via example sheets.

Lecturer: Stefan Adams (Stefan Grosskinsky; Roman Kotecký; Roger Tribe; Oleg Zaboronski; Agelos Georgakopoulos)

Who is This Module For?

This is a lecture course in probability in high dimensions with a view toward applications in data sciences and machine learning. In its elementary version it is intended for third year students who like to get some overview on concentration of measure problems in high dimensions and basic machine learning. Measure concentration ideas developed during the last century in various parts of mathematics including functional analysis, probability theory, and statistical mechanics, areas typically dealing with models involving an infinite number of variables.

Why this module?

The data sciences are moving fast, and probabilistic methods often provide a foundation and inspiration for such advances. Today, a typical probability course is no longer sufficient to acquire the level of mathematical sophistication that is expected from a beginning researcher or employee in applied mathematics, data science or machine learning. The module is intended to partially cover this gap. It presents some key probabilistic methods and results that form an essential toolbox for a mathematical data scientist.

What is this module about?

High-dimensional probability is an area of probability theory that studies random objects in $\mathbb{R}^n$, where the dimension n can be very large. The module places particular emphasis on random vectors, random matrices, and random projections. It teaches basic theoretical skills for the analysis of these objects, which include concentration inequalities, covering and packing arguments, decoupling and symmetrisation tricks, chaining and comparison techniques for stochastic processes, combinatorial reasoning based on the dimension, and a lot more. The study of high-dimensional probability provides vital theoretical tools for applications in data science. The book integrates theory with applications for covariance estimation, semidefinite programming, networks, elements of statistical learning, error correcting codes, clustering, matrix completion, dimension reduction, and sparse signal recovery, statistical mechanics and machine learning.