Chair
• Dan Liviu Nicolae, Statistics and Medicine

Professors
• Yali Amit, Statistics and Computer Science
• Mihai Anitescu, Statistics and Argonne National Laboratory
• Guillaume Bal, Statistics and Mathematics
• Rina Foygel Barber
• Brent Doiron, Neurobiology and Statistics
• Lars Peter Hansen, Economics and Statistics
• Steven Lalley
• Gregory F. Lawler, Mathematics and Statistics
• Lek-Heng Lim
• Peter McCullagh
• Mary Sara McPeek, Statistics and Human Genetics
• Per Mykland, Statistics and the Stevanovich Center
• Dan Liviu Nicolae, Statistics and Medicine
• John Reinitz, Statistics, Ecology and Evolution, and Molecular Genetics and Cell Biology
• Mary Silber
• Matthew Stephens, Statistics and Human Genetics
• Stephen M. Stigler
• Rebecca Willett, Statistics and Computer Science
• Kirk M. Wolter
• Wei Biao Wu

Associate Professors
• Imre Risi Kondor, Computer Science and Statistics

Assistant Professors
• Claire Donnat
• Chao Gao
• Jeremy Hoskins
• Yuehaw Khoo
• Alisa Knizel
• Daniel Sanz-Alonso
• Yi Sun
• Victor Veitch
• Jingshu Wang

Senior Instructional Professors
• Eric Baer
• David Biron
• Mei Wang

Instructional Professors
• Kendra Burbank
• Yibi Huang
• Fei Liu

Instructors
• Daniel Massatt
• Bradley J. Nelson
• Benjamin Palacios
The Department of Statistics offers an exciting and revamped graduate program that prepares students for cutting-edge interdisciplinary research in a wide variety of fields. The field of statistics has become a core component of research in the biological, physical, and social sciences, as well as in traditional computer science domains such as artificial intelligence. In light of this, the Department of Statistics is currently undergoing a major expansion of approximately ten new faculty into fields of Computational and Applied Mathematics. The massive increase in the data acquired, through scientific measurement on one hand and through web-based collection on the other, makes the development of statistical analysis and prediction methodologies more relevant than ever. Our graduate program aims to prepare students to address these issues through rigorous training in theory, methodology, and applications of statistics; rigorous training in scientific computation; and research projects in core methodology of statistics and computation as well as in a wide variety of interdisciplinary fields.

The Department of Statistics offers two tracks of graduate study, one leading to the Master of Science (M.S.) degree, the other to the Doctorate of Philosophy (Ph.D.). The M.S. degree is a professional degree. Students who receive this degree are prepared for nonacademic careers in which the use of advanced statistical and computational methods is of central importance. The program also prepares students for possible further graduate study.

During the first year of the Ph.D. program, students are given a thorough grounding in material that forms the foundations of modern statistics and scientific computation, including data analysis, mathematical statistics, probability theory, applied probability and modeling, and computational methods. Throughout the entire program, students attend a weekly consulting seminar where researchers from across the University come to get advice on modeling, statistical analysis, and computation. This seminar is often the source of interesting and ongoing research projects.

In the second year, students have a wide range of choices of topics they can pursue further, based on their interests, through advanced courses and reading courses with faculty. During the second year, students will typically identify their subfield of interest, take some advanced courses in the subject, and interact with the relevant faculty members. The Department maintains very strong connections to numerous other units on campus, either through joint appointments of the faculty or through ongoing collaborations. Students have easy access to faculty in other departments, which allows them to expand their interactions and develop new interdisciplinary research projects. Examples include joint projects with Human Genetics, Ecology and Evolution, Neurobiology, Chemistry, Economics, Health Studies, and Astronomy.

The program offers four core sequences:

- Probability (STAT 30400, 38100, 38300)
- Mathematical statistics (STAT 30400, 30100, 30210)
- Applied statistics (STAT 34300, 34700, 34800)
- Computational mathematics and machine learning (STAT 30900, 31015/31020, 37710)

All students must take the applied statistics sequence and one of the two theoretical sequences: mathematical statistics or probability. At the start of their second year, the students take preliminary examinations covering the two sequences they have taken. In addition, it is highly recommended that students take a third core sequence based on their interests and in consultation with the Department Graduate Adviser (DGA). Incoming first-year students may request the DGA to take one or both of their preliminary exams. If approved, and if the student passes one or more of these, the student will then be excused from the requirement of taking the first-year courses in that subject. Incoming students are advised by the DGA until they find a faculty adviser for their Ph.D. thesis work.

In their second year, Ph.D. students typically take a number of advanced topics: courses in statistics, probability, computation, and applications. These should be selected with the dual objective of (i) acquiring a broad overview of current research areas, and (ii) settling on a particular research topic and dissertation supervisor. It is recommended that the students take at least one regular class based course each quarter. In addition, students can ask to take reading courses with faculty to learn more in depth about their fields of research. Students have considerable latitude in selecting their second-year courses, but their programs must be approved by the Department Graduate Adviser. Students are expected to find a dissertation adviser by the end of the second year. The detailed process is listed here (https://stat.uchicago.edu/admissions/about-phd/).

**The Ph.D.: Training in Teaching, Presentation, and Consulting**

Part of every statistician’s job is to evaluate the work of others and to communicate knowledge, experience, and insights. Every statistician is, to some extent, an educator, and the department provides graduate students with training for this aspect of their professional lives. The department expects all doctoral students, regardless
of their professional objectives and sources of financial support, to take part in a graduated program of participation in some or all phases of instruction, from grading, course assisting, and conducting discussion sections, to being a lecturer with responsibility for an entire course.

Students also receive training in how to present research in short seminars in the first and second years of study. Later, students present their own work in a dissertation proposal and, eventually, in a thesis defense. The student seminars are listed here (https://stat.uchicago.edu/events/).

Ph.D. students should also participate in the department’s consulting program (https://stat.uchicago.edu/about/consulting/), which is led by faculty members and exposes the students to empirical projects inside the university. Projects are carried out by groups of students under the guidance of a faculty member. The client is a researcher in an applied area, usually associated with the university. An informal seminar meets regularly over lunch to provide a forum for presenting and discussing problems, solutions, and topics in statistical consultation. Students present interesting or difficult consulting problems to the seminar as a way of stimulating wider consideration of the problem and as a means of developing familiarity with the kinds of problems and lines of attack involved. Often the client will participate in the presentation and discussion.

PROGRAMS AND REQUIREMENTS FOR THE M.S. DEGREE

The main requirements of the M.S. program are a sequence of at least nine approved courses plus a Master’s paper. Students may take up to two years of courses. A detailed set of regulations can be found here (https://stat.uchicago.edu/admissions/about-ms/). (http://www.stat.uchicago.edu/students/master.shtml/) A substantial fraction of available courses are the same as for the Ph.D. degree.

FACILITIES

Almost all departmental activities—classes, seminars (https://stat.uchicago.edu/events/statistics-colloquium/), computation (https://www.stat.uchicago.edu/intranet/computing/index.shtml/), and student and faculty offices (https://stat.uchicago.edu/people/)—are located in Jones Laboratory (https://d3q0k55sx5f.cloudfront.net/stat/docs/People/UChicago_Maps--Jones.pdf?cachebuster:32). Each student is assigned a desk in one of several offices. The major computing facilities of the department are based upon a network of PCs running mainly Linux. One computer room currently houses many of these PCs; these rooms are directly and primarily for graduate students in the Statistics Department. In addition, all student offices have limited computer facilities. For further information, consult the department’s computing policies (https://www.stat.uchicago.edu/intranet/computing/policies/index.shtml/).

STATISTICS THROUGHOUT THE UNIVERSITY

In addition to the courses, seminars, and programs in the Department of Statistics, courses and workshops of direct interest to statisticians occur throughout the University, most notably in the programs in statistics and econometrics in the Booth School of Business (http://www.chicagobooth.edu/) and in the research programs in Public Health Sciences (http://health.bsd.uchicago.edu/), Human Genetics (http://genes.uchicago.edu/), Financial Mathematics and Econometrics (https://finmath.uchicago.edu/), Computer Science (http://www.cs.uchicago.edu/), Economics (http://economics.uchicago.edu/), and NORC (http://www.norc.uchicago.edu/) (formerly the National Opinion Research Center). The large number of statistics related seminars (https://stat.uchicago.edu/events/) is perhaps the best indication of the vibrancy of the statistics research community here at the University of Chicago.

STATISTICS COURSES

STAT 30030. Statistical Theory and Methods Ia. 100 Units.

This course is the first quarter of a two-quarter sequence providing a principled development of statistical methods, including practical considerations in applying these methods to the analysis of data. The course begins with a brief review of probability and some elementary stochastic processes, such as Poisson processes, that are relevant to statistical applications. The bulk of the quarter covers principles of statistical inference from both frequentist and Bayesian points of view. Specific topics include maximum likelihood estimation, posterior distributions, confidence and credible intervals, principles of hypothesis testing, likelihood ratio tests, multinomial distributions, and chi-square tests. Additional topics may include diagnostic plots, bootstrapping, a critical comparison of Bayesian and frequentist inference, and the role of conditioning in statistical inference. Examples are drawn from the social, physical, and biological sciences. The statistical software package R will be used to analyze datasets from these fields and instruction in the use of R is part of the course.

Instructor(s): Staff Terms Offered: Autumn

Prerequisite(s): STAT 25100 or STAT 25150 or MATH 23500. This course is only open to graduate students in Statistics, Applied Mathematics, and Financial Mathematics, and to undergraduate Statistics majors, or by consent of instructor.

Note(s): Some previous experience with statistics helpful but not required. Concurrent or prior linear algebra (MATH 19620 or 20250 or STAT 24300 or equivalent) is recommended for students continuing to STAT 24510. Students may count either STAT 24400 or STAT 24410, but not both, toward the forty-two credits required for graduation.

Equivalent Course(s): STAT 24410
STAT 30400. Distribution Theory. 100 Units.
This course is a systematic introduction to random variables and probability distributions. Topics include standard distributions (i.e., uniform, normal, beta, gamma, $F$, $t$, Cauchy, Poisson, binomial, and hypergeometric); properties of the multivariate normal distribution and joint distributions of quadratic forms of multivariate normal; moments and cumulants; characteristic functions; exponential families; modes of convergence; central limit theorem; and other asymptotic approximations.

Instructor(s): Staff Terms Offered: Autumn
Prerequisite(s): STAT 24500 or STAT 30100
Equivalent Course(s): STAT 24510

STAT 30600. Adv. Statistical Inference 1. 100 Units.
This course continues the development of Mathematical Statistics, with an emphasis on hypothesis testing. Topics include comparison of Bayesian and frequentist hypothesis testing; admissibility of Bayes' rules; confidence and credible sets; likelihood ratio tests and their asymptotics; Bayes factors; methods for assessing predictions for normal means; shrinkage and thresholding methods; sparsity; shrinkage as an example of empirical Bayes; multiple testing and false discovery rates; Bayesian approach to multiple testing; sparse linear regressions (subset selection and LASSO, proof of estimation errors for LASSO, Bayesian perspective of sparse regressions); and Bayesian model averaging.

Instructor(s): Staff Terms Offered: Winter
Prerequisite(s): STAT 30400 or consent of instructor

STAT 30750. Numerical Linear Algebra. 100 Units.
This course is devoted to the basic theory of linear algebra and its significant applications in scientific computing. The objective is to provide a working knowledge and hands-on experience of the subject suitable for graduate level work in statistics, econometrics, quantum mechanics, and numerical methods in scientific computing. Topics include Gaussian elimination, vector spaces, linear transformations and associated fundamental subspaces, orthogonality and projections, eigenvectors and eigenvalues, diagonalization of real symmetric and complex Hermitian matrices, the spectral theorem, and matrix decompositions (QR, Cholesky and Singular Value Decompositions). Systematic methods applicable in high dimensions and techniques commonly used in scientific computing are emphasized. Students enrolled in the graduate level STAT 30750 will have additional work in assignments, exams, and projects including applications of matrix algebra in statistics and numerical
computations implemented in Matlab or R. Some programming exercises will appear as optional work for students enrolled in the undergraduate level STAT 24300.

Terms Offered: Autumn
Prerequisite(s): Multivariate calculus (MATH 15910 or MATH 16300 or MATH 16310 or MATH 18400 or MATH 19520 or MATH 20000 or MATH 20510 or MATH 20900 or PHYS 22100 or equivalent). Previous exposure to linear algebra is helpful.
Equivalent Course(s): STAT 24300

STAT 30800. Advanced Statistical Inference II. 100 Units.
This course will discuss the following topics in high-dimensional statistical inference: random matrix theory and asymptotics of its eigen-decompositions, estimation and inference of high-dimensional covariance matrices, large dimensional factor models, multiple testing and false discovery control and high-dimensional semiparametrics. On the methodological side, probability inequalities, including exponential, Nagaev, and Rosenthal-type inequalities will be introduced.
Terms Offered: To be determined; may not be offered in 2020-2021.
Prerequisite(s): STAT 30400, STAT 30100, and STAT 30210, or consent of instructor

STAT 30810. High Dimensional Time Series Analysis. 100 Units.
This course will include lectures on the following topics: review of asymptotics for low dimensional time series analysis (linear and nonlinear processes; nonparametric methods; spectral and time domain approaches); covariance, precision, and spectral density matrix estimation for high dimensional time series; factor models; estimation of high dimensional vector autoregressive processes; prediction; and high dimensional central limit theorems under dependence.
Terms Offered: To be determined

STAT 30850. Multiple Testing, Modern Inference, and Replicability. 100 Units.
This course examines the problems of multiple testing and statistical inference from a modern point of view. High-dimensional data is now common in many applications across the biological, physical, and social sciences. With this increased capacity to generate and analyze data, classical statistical methods may no longer ensure the reliability or replicability of scientific discoveries. We will examine a range of modern methods that provide statistical inference tools in the context of modern large-scale data analysis. The course will have weekly assignments as well as a final project, both of which will include both theoretical and computational components.
Terms Offered: To be determined
Prerequisite(s): STAT 24400 or STAT 24410. Familiarity with regression and with coding in R are recommended.
Equivalent Course(s): STAT 27850

STAT 30900. Mathematical Computation I: Matrix Computation Course. 100 Units.
This is an introductory course on numerical linear algebra, which is quite different from linear algebra. We will be much less interested in algebraic results that follow from axiomatic definitions of fields and vector spaces but much more interested in analytic results that hold only over the real and complex fields. The main objects of interest are real- or complex-valued matrices, which may come from differential operators, integral transforms, bilinear and quadratic forms, boundary and coboundary maps, Markov chains, correlations, DNA microarray measurements, movie ratings by viewers, friendship relations in social networks, etc. Numerical linear algebra provides the mathematical and algorithmic tools for analyzing these matrices. Topics covered: basic matrix decompositions LU, QR, SVD; Gaussian elimination and LU/LDU decompositions; backward error analysis, Gram-Schmidt orthogonalization and QR/complete orthogonal decompositions; solving linear systems, least squares, and total least squares problem; low-rank matrix approximations and matrix completion. We shall also include a brief overview of stationary and Krylov subspace iterative methods; eigenvalue and singular value problems; and sparse linear algebra.
Terms Offered: Autumn
Prerequisite(s): Linear algebra (STAT 24300 or equivalent) and some previous experience with statistics.
Equivalent Course(s): CAAM 30900, CMSC 37810

STAT 31015. Mathematical Computation IIA: Convex Optimization. 100 Units.
The course will cover techniques in unconstrained and constrained convex optimization and a practical introduction to convex duality. The course will focus on (1) formulating and understanding convex optimization problems and studying their properties; (2) understanding and using the dual; and (3) presenting and understanding optimization approaches, including interior point methods and first order methods for nonsmooth problems. Examples will be mostly from data fitting, statistics and machine learning.
Instructor(s): Nathan Srebro Terms Offered: Winter
Prerequisite(s): STAT 30900/CMSC 37810
Equivalent Course(s): CAAM 31015, BUSN 36903, TTIC 31070, CMSC 35470

STAT 31020. Mathematical Computation IIB: Nonlinear Optimization. 100 Units.
This course covers the fundamentals of continuous optimization with an emphasis on algorithmic and computational issues. The course starts with the study of optimality conditions and techniques for unconstrained optimization, covering line search and trust region approaches, and addressing both factorization-based and iterative methods for solving the subproblems. The Karush-Kuhn-Tucker conditions for general constrained and nonconvex optimization are then discussed and used to define algorithms for constrained optimization including augmented Lagrangian, interior-point and (if time permits) sequential quadratic programming.
Iterative methods for large sparse problems, with an emphasis on projected gradient methods, will be presented. Several substantial programming projects (using MATLAB and aiming at both data-intensive and physical sciences applications) are completed during the course.

Terms Offered: Winter
Prerequisite(s): STAT 30900/CMSC 37810
Equivalent Course(s): CAAM 31020

STAT 31080. Numerical Analysis for Statistics and Applied Mathematics. 100 Units.
This is a beginning graduate course on selected numerical methods used in modern statistics and applied mathematics. Topics include fundamentals of ODEs and PDEs, quadratures, and Monte Carlo methods. Methods of analysis are introduced including error measures and different notions of numerical convergence. Newton’s method, convex optimization and elements of nonconvex optimization are covered, together with implementations in selected selected software packages.

Terms Offered: To be determined
Prerequisite(s): STAT 24300 or background in linear algebra.

STAT 31100. Mathematical Computation III: Numerical Methods for PDE’s. 100 Units.
The first part of this course introduces basic properties of PDE’s; finite difference discretizations; and stability, consistency, convergence, and Lax’s equivalence theorem. We also cover examples of finite difference schemes; simple stability analysis; convergence analysis and order of accuracy; consistency analysis and errors (i.e., dissipative and dispersive errors); and unconditional stability and implicit schemes. The second part of this course includes solution of stiff systems in 1, 2, and 3D; direct vs. iterative methods (i.e., banded and sparse LU factorizations); and Jacobi, Gauss-Seidel, multigrid, conjugate gradient, and GMRES iterations.

Terms Offered: Spring
Prerequisite(s): Some prior exposure to differential equations and linear algebra
Equivalent Course(s): MATH 38309, CMSC 37812, CAAM 31100

STAT 31120. Numerical Methods for Stochastic Differential Equations. 100 Units.
The numerical analysis of SDE differs significantly from that of ODE due to the peculiarities of stochastic calculus. This course starts with a brief review of stochastic calculus and stochastic differential equations, then emphasizing the numerical methods needed to solve such equations. The stochastic Taylor expansion provides the basis for the discrete-time numerical methods for differential equations. The course presents many results on high-order methods for strong sample path approximations and for weak functional approximations. To help with developing an intuitive understanding of the underlying mathematics and hand-on numerical skills, examples and exercises on PC are included.

Terms Offered: Spring
Prerequisite(s): Knowledge of ODE and SDE is essential. STAT 39000 or STAT 39010 or STAT 38510 are strongly recommended.
Equivalent Course(s): CAAM 31120

STAT 31140. Computational Imaging: Theory and Methods. 100 Units.
Computational imaging refers to the process of forming images from data where computation plays an integral role. This course will cover basic principles of computational imaging, including image denoising, regularization techniques, linear inverse problems and optimization-based solvers, and data acquisition models associated with tomography and interferometry. Specific topics may include patch-based denoising, sparse coding, total variation, dictionary learning, computational photography, compressive imaging, inpainting, and deep learning for image reconstruction.

Instructor(s): R. Willett Terms Offered: To be determined
Equivalent Course(s): CMSC 31140, CAAM 31140

STAT 31150. Inverse Problems and Data Assimilation. 100 Units.
This class provides an introduction to Bayesian Inverse Problems and Data Assimilation, emphasizing the theoretical and algorithmic inter-relations between both subjects. We will study Gaussian approximations and optimization and sampling algorithms, including a variety of Kalman-based and particle filters as well as Markov chain Monte Carlo schemes designed for high-dimensional inverse problems.

Instructor(s): D. Sanz-Alonso Terms Offered: Autumn
Prerequisite(s): Familiarity with calculus, linear algebra, and probability/statistics at the level of STAT 24400 or STAT 24410. Some knowledge of ODEs may also be helpful.
Equivalent Course(s): CAAM 31150

STAT 31190. Fast Algorithms. 100 Units.
This course will introduce students to several classes of computational methods broadly referred to as “fast analysis-based algorithms” which exploit information about structure and symmetry to obtain more favorable computational complexity. Examples which will be discussed are butterfly algorithms, fast multipole methods, fast direct solvers, and hierarchical matrix compression. Though many of these algorithms first arose in physical applications such as simulating the motion of stars or the propagation of light and sound, they have subsequently found many fruitful applications in signal processing and data science.

Terms Offered: Spring
Prerequisite(s): Familiarity with PDEs, analysis, and programming.
STAT 31200. Introduction to Stochastic Processes I. 100 Units.
This course introduces stochastic processes not requiring measure theory. Topics include branching processes, recurrent events, renewal theory, random walks, Markov chains, Poisson, and birth-and-death processes.
Instructor(s): Staff Terms Offered: To be determined
Prerequisite(s): STAT 25100 and MATH 20500; STAT 30400 or consent of instructor
Note(s): Students with credit for MATH 235 should not enroll in STAT 312.
STAT 31210. Applied Functional Analysis. 100 Units.
This course will cover classical topics of applied functional analysis: description of functional spaces such as Banach spaces and Hilbert spaces; properties of linear operators acting on such spaces, compactness and spectral decomposition of compact operators; and applications to ordinary and partial differential equations.
Terms Offered: To be determined
Equivalent Course(s): CAAM 31210
STAT 31220. Partial Differential Equations. 100 Units.
This is an introduction to the theory of partial differential equations covering representation formulas and regularity theory for elliptic, parabolic, and hyperbolic equations; the method of characteristics; variational formulations for second-order linear elliptic equations; and the calculus of variations.
Terms Offered: Winter
Equivalent Course(s): CAAM 31220
STAT 31230. Inverse Problems in Imaging. 100 Units.
This course focuses on the mathematical description of many inverse problems that appear in geophysical and medical imaging: X-ray tomography, ultrasound tomography and seismic imaging, optical and electrical tomography, as well as more recent imaging modalities such as elastography and photo-acoustic tomography. Seen as reconstructions of constitutive parameters in differential equations from redundant boundary measurements, these continuous models tell us which parameters may or may not be reconstructed, and with which stability with respect to measurement errors. Time-permitting, we will also consider general methodologies to perform such reconstructions (regularization, optimization, Bayesian framework). Some knowledge of PDE and Fourier transforms is recommended.
Terms Offered: Spring
Prerequisite(s): STAT 31220
Equivalent Course(s): CAAM 31230
STAT 31240. Variational Methods in Image Processing. 100 Units.
This course discusses mathematical models arising in image processing. Topics covered will include an overview of tools from the calculus of variations and partial differential equations, applications to the design of numerical methods for image denoising, deblurring, and segmentation, and the study of convergence properties of the associated models. Students will gain an exposure to the theoretical basis for these methods as well as their practical application in numerical computations.
Terms Offered: Spring
Equivalent Course(s): CAAM 31240
STAT 31410. Applied Dynamical Systems. 100 Units.
This course is an introduction to dynamical systems for analysis of nonlinear ordinary differential equations. The focus is on methods of bifurcation theory, canonical examples of forced nonlinear oscillators, fast-slow systems, and chaos. Examples will be drawn from mathematical modeling of physical and biological systems. While geometric perspectives will be emphasized, assignments will also introduce asymptotic methods for analysis and use numerical simulation as an exploratory tool. This course assumes students have a background in ordinary differential equations and linear algebra at the undergraduate level and an interest in mathematical modeling for applications.
Instructor(s): M. Silber Terms Offered: Spring
Prerequisite(s): ODEs and/or dynamical systems at an undergraduate level or consent of instructor.
Equivalent Course(s): CAAM 31410
STAT 31430. Applied Linear Algebra. 100 Units.
This course will provide a review and development of topics in linear algebra aimed toward preparing students for further graduate coursework in Computational and Applied Mathematics. Topics will include discussion of matrix factorizations (including diagonalization, the spectral theorem for normal matrices, the singular value decomposition, and the Schur and polar decompositions), and an overview of classical direct and iterative approaches to numerical methods for problems formulated in the language of linear algebra (including the conjugate gradient method). Additional topics will be included depending on student interests.
Instructor(s): E. Baer Terms Offered: Autumn
Prerequisite(s): STAT 24300 or MATH 20250 or Graduate Student in Physical Sciences Division
Equivalent Course(s): CAAM 31430
STAT 31440. Applied Analysis. 100 Units.
This course provides an overview of fundamentals of mathematical analysis with an eye towards developing the toolkit of graduate students in applied mathematics. Topics covered include metric spaces and basic topological notions, aspects of mathematical analysis in several variables, and an introduction to measure and integration.
Instructor(s): E. Baer
Terms Offered: Autumn
Equivalent Course(s): CAAM 31440

STAT 31450. Applied Partial Differential Equations. 100 Units.
Partial differential equations (PDEs) are used to model applications in a wide variety of fields: fluid dynamics, optics, atomic and plasma physics, elasticity, chemical reactions, climate modeling, stock markets, etc. The study of their mathematical structure and solution methods remains at the forefront of applied mathematics. The course concentrates on deriving an important set of examples of PDEs from simple physical models, which are often closely related to those describing more complex physical systems. The course will also cover analytical methods and tools for solving these PDEs; such as separation of variables, Fourier series and transforms, Sturm-Liouville theory, and Green's functions. The course is suitable for graduate students and advanced undergraduates in science, engineering, and applied mathematics.
Terms Offered: Spring
Prerequisite(s): Instructor consent.
Equivalent Course(s): CAAM 31450

STAT 31460. Applied Fourier Analysis. 100 Units.
Decompositions of functions into frequency components via the Fourier transform, and related sparse representations, are fundamental tools in applied mathematics. These ideas have been important in applications to signal processing, imaging, and the quantitative and qualitative analysis of a broad range of mathematical models of data (including modern approaches to machine learning) and physical systems. Topics to be covered in this course include an overview of classical ideas related to Fourier series and the Fourier transform, wavelet representations of functions and the framework of multiresolution analysis, and applications throughout computational and applied mathematics.
Terms Offered: Winter
Prerequisite(s): Graduate student in the Physical Sciences Division or consent of instructor.
Equivalent Course(s): CAAM 31460

STAT 31511. Monte Carlo Simulation. 100 Units.
This class primarily concerns the design and analysis of Monte Carlo sampling techniques for the estimation of averages with respect to high dimensional probability distributions. Standard simulation tools such as importance sampling, Metropolis-Hastings, Langevin dynamics, and hybrid Monte Carlo will be introduced along with basic theoretical concepts regarding their convergence to equilibrium. The class will explore applications of these methods in Bayesian statistics and machine learning as well as to other simulation problems arising in the physical and biological sciences. Particular attention will be paid to the major complicating issues like conditioning (with analogies to optimization) and rare events and methods to address them.
Terms Offered: To be determined
Prerequisite(s): Multivariate calculus and linear algebra; elementary knowledge of ordinary differential equations.
Equivalent Course(s): CAAM 31511

STAT 31521. Applied Stochastic Processes. 100 Units.
This course concerns the estimation of the dynamic properties of time-dependent stochastic systems. The class will begin with an introduction to the numerical simulation of continuous time Markov processes including the discretization of stochastic (and ordinary) differential equations. Problems associated with multiple time scales will be discussed along with methods to address them (implicit discretizations, multiscale methods and dimensional reduction). The class will also cover interacting particle methods and other techniques for the efficient simulation of dynamical rare events.
Terms Offered: Winter. To be determined
Prerequisite(s): Multivariate calculus and linear algebra

STAT 31550. Uncertainty Quantification. 100 Units.
This course will cover mathematical, statistical, and algorithmic questions that arise at the interface of complex modeling and data processing. Emphasis will be given to characterizing and quantifying the uncertainties inherent in the use of models and to exploring principled ways to reduce said uncertainty by the use of data. Specific topics include Bayesian inverse problems and data assimilation.
Terms Offered: To be determined
Prerequisite(s): STAT 30200 or consent of instructor

STAT 31610. Mathematical Aspects of Electronic Structure of Materials. 100 Units.
This course considers mathematical and numerical methods to approach electronic structure of materials through several hot-topic examples including topological insulators and incommensurate 2D materials in addition to classical systems such as periodic crystals. The course will begin with a discussion of the basics of quantum mechanics for those not yet familiar before moving to models designed for varying system sizes, from DFT to tight-binding. The theory and numerical tools for studying observables such as Chern numbers, conductivity, and density of states will be considered.
Equivalent Course(s): CAAM 31610

STAT 31700. Introduction to Probability Models. 100 Units.
This course introduces stochastic processes as models for a variety of phenomena in the physical and biological sciences. Following a brief review of basic concepts in probability, we introduce stochastic processes that are popular in applications in sciences (e.g., discrete time Markov chain, the Poisson process, continuous time Markov process, renewal process and Brownian motion).
Instructor(s): Staff Terms Offered: To be determined
Prerequisite(s): STAT 24400 or STAT 24410 or STAT 25100 or STAT 25150
Equivalent Course(s): STAT 25300

STAT 31900. Introduction to Causal Inference. 100 Units.
This course is designed for graduate students and advanced undergraduate students from the social sciences, education, public health science, public policy, social service administration, and statistics who are involved in quantitative research and are interested in studying causality. The goal of this course is to equip students with basic knowledge of and analytic skills in causal inference. Topics for the course will include the potential outcomes framework for causal inference; experimental and observational studies; identification assumptions for causal parameters; potential pitfalls of using ANCOVA to estimate a causal effect; propensity score based methods including matching, stratification, inverse-probability-of-treatment-weighting (IPTW), marginal mean weighting through stratification (MMWS), and doubly robust estimation; the instrumental variable (IV) method; regression discontinuity design (RDD) including sharp RDD and fuzzy RDD; difference in difference (DID) and generalized DID methods for cross-section and panel data, and fixed effects model. Intermediate Statistics or equivalent such as STAT 224/PHS 324, PP 31301, BUS 41100, or SOC 30005 is a prerequisite. This course is a prerequisite for "Advanced Topics in Causal Inference" and "Mediation, moderation, and spillover effects."
Instructor(s): G. Hong Terms Offered: This course is not offered in 2021-22.
Prerequisite(s): Intermediate Statistics or equivalent such as STAT 224/PHS 324, PP 31301, BUS 41100, or SOC 30005
Note(s): CHDV Distribution: M; M This course will not be offered in the academic year 2021-22.
Equivalent Course(s): PHS 43201, MACS 51000, PLSC 30102, SOCI 30315, CHDV 30102

STAT 32400. Probability and Statistics. 100 Units.
This Ph.D.-level course (in addition to BUSF 41902/STAT 32500) provides a thorough introduction to Classical and Bayesian statistical theory. The two quarter sequence provides the necessary probability and statistical background for many of the advanced courses in the Chicago Booth curriculum. The central topic is probability. Basic concepts in probability are covered. An introduction to martingales is given. Homework assignments are given throughout the quarter. Course description is subject to change. Please visit the Booth portal and search via the course search tool for the most up to date information: http://boothportal.chicagobooth.edu/portal/server.pt/community/course_search
Terms Offered: Autumn
Equivalent Course(s): BUSN 41901

STAT 32900. Applied Multivariate Analysis. 100 Units.
The course will introduce the basic theory and applications for analyzing multidimensional data. Topics include multivariate distributions, Gaussian models, multivariate statistical inferences and applications, classifications, cluster analysis, and dimension reduction methods. Course content is subject to change in order to keep the contents up-to-date with new development in multivariate statistical techniques. The course is offered in alternate years by the Statistics Department (S15, S17, ... ) and the Booth Business School (S16, S18, ... ). When the course is offered by the Booth school, please visit the Booth portal and search via the course search tool: http://boothportal.chicagobooth.edu/portal/server.pt/community/course_search for the most up to date information.
Equivalent Course(s): BUSN 41912

STAT 32940. Multivariate Data Analysis via Matrix Decompositions. 100 Units.
This course is about using matrix computations to infer useful information from observed data. One may view it as an “applied” version of Stat 30900 although it is not necessary to have taken Stat 30900; the only prerequisite for this course is basic linear algebra. The data analytic tools that we will study will go beyond linear and multiple regression and often fall under the heading of "Multivariate Analysis" in Statistics. These include factor analysis, correspondence analysis, principal components analysis, multidimensional scaling, discriminant analysis, clustering, and correspondence analysis (if time permits). Further topics on statistical learning for high dimensional data and complex structures include penalized regression models (LASSO, ridge, elastic net), sparse PCA, independent component analysis, Gaussian mixture model, Expectation-Maximization methods, and
random forest. Theoretical derivations will be presented with emphasis on motivations, applications, and hands-on data analysis.

Instructor(s): M. Wang Terms Offered: Spring
Prerequisite(s): (STAT 24300 or MATH 20250) and (STAT 24500 or STAT 24510). Graduate students in Statistics or Financial Mathematics can enroll without prerequisites.
Note(s): Linear algebra at the level of STAT 24300. Knowledge of probability and statistical estimation techniques (e.g. maximum likelihood and linear regression) at the level of STAT 24400-24500.
Equivalent Course(s): STAT 24620

STAT 33100. Sample Surveys. 100 Units.
This course covers random sampling methods; stratification, cluster sampling, and ratio estimation; and methods for dealing with nonresponse and partial response.
Instructor(s): K. Wolter Terms Offered: Autumn
Prerequisite(s): Consent of instructor

STAT 33500. Time-series Analysis for Forecasting and Model Building. 100 Units.
Forecasting plays an important role in business planning and decisionmaking. This Ph.D.-level course discusses time series models that have been widely used in business and economic data analysis and forecasting. Both theory and methods of the models are discussed. Real examples are used throughout the course to illustrate applications. The topics covered include: (1) stationary and unit-root non-stationary processes; (2) linear dynamic models, including Autoregressive Moving Average models; (3) model building and data analysis; (4) prediction and forecasting evaluation; (5) asymptotic theory for estimation including unit-root theory; (6) models for time varying volatility; (7) models for time varying correlation including Dynamic Conditional Correlation and time varying factor models.; (9) state-space models and Kalman filter; and (10) models for high frequency data. Course description is subject to change. Please visit the Booth portal and search via the course search tool for the most up to date information: http://boothportal.chicagobooth.edu/portal/server.pt/community/course_search/
Prerequisite(s): BUSF 41901/STAT 32400 or instructor consent.
Equivalent Course(s): BUSN 41910

STAT 33600. Time Dependent Data. 100 Units.
This course considers the modeling and analysis of data that are ordered in time. The main focus is on quantitative observations taken at evenly spaced intervals and includes both time-domain and spectral approaches.
Instructor(s): W. Wu
Prerequisite(s): STAT 24500 w/B- or better or STAT 24510 w/C+ or better is required; alternatively STAT 22400 w/B- or better and exposure to multivariate calculus (MATH 16300 or MATH 16310 or MATH 18400 or MATH 19520 or MATH 20000 or MATH 20500 or MATH 20510 or MATH 20800). Graduate students in Statistics or Financial Mathematics can enroll without prerequisites. Some previous exposure to Fourier series is helpful but not required.
Equivalent Course(s): STAT 26100

STAT 33700. Multivariate Time Series Analysis. 100 Units.
This course investigates the dynamic relationships between variables. It starts with linear relationships between two variables, including distributed-lag models and detection of unidirectional dependence (Granger causality). Nonlinear and time-varying relationships are also discussed. Dynamic models discussed include vector autoregressive models, vector autoregressive moving-average models, multivariate regression models with time series errors, co-integration and error-correction models, state-space models, dynamic factor models, and multivariate volatility models such as BEKK, Dynamic conditional correlation, and copula-based models. The course also addresses impulse response function, structural specification, co-integration tests, least squares estimates, maximum likelihood estimates, principal component analysis, asymptotic principal component analysis, principal volatility components, recursive estimation, and Markov Chain Monte Carlo estimation. Empirical data analysis is an integral part of the course. Students are expected to analyze many real data sets. The main software used in the course is the MTS package in R, but students may use their own software if preferred.
Equivalent Course(s): BUSN 41914

STAT 33910. Financial Statistics: Time Series, Forecasting, Mean Reversion, and High Frequency Data. 100 Units.
This course is an introduction to the econometric analysis of high-frequency financial data. This is where the stochastic models of quantitative finance meet the reality of how the process really evolves. The course is focused on the statistical theory of how to connect the two, but there will also be some data analysis. With some additional statistical background (which can be acquired after the course), the participants will be able to read articles in the area. The statistical theory is longitudinal, and it thus complements cross-sectional calibration methods (implied volatility, etc.). The course also discusses volatility clustering and market microstructure.
Terms Offered: Winter
Prerequisite(s): Some statistics/econometrics background as in STAT 24400-24500, or FINM 33150 and FINM 33400, or equivalent, or consent of instructor.
Equivalent Course(s): FINM 33170
STAT 34300. Applied Linear Stat Methods. 100 Units.
This course introduces the theory, methods, and applications of fitting and interpreting multiple regression models. Topics include the examination of residuals, the transformation of data, strategies and criteria for the selection of a regression equation, nonlinear models, biases due to excluded variables and measurement error, and the use and interpretation of computer package regression programs. The theoretical basis of the methods, the relation to linear algebra, and the effects of violations of assumptions are studied. Techniques discussed are illustrated by examples involving both physical and social sciences data.
Terms Offered: Autumn
Prerequisite(s): Graduate student in Statistics or Financial Mathematics or instructor consent.
Note(s): Students who need it should take Linear Algebra (STAT 24300 or equivalent) concurrently.

STAT 34700. Generalized Linear Models. 100 Units.
This applied statistics course is a successor of STAT 34300 and covers the foundations of generalized linear models (GLM). We will discuss the general linear modeling idea for exponential family data and introduce specifically models for binary, multinomial, count and categorical data, and the challenges in model fitting, and inference. We will also discuss approaches that supplement the classical GLM, including quasi-likelihood for over-dispersed data, robust estimation, and penalized GLM. The course also covers related topics including mixed effect models for clustered data, the Bayesian approach of GLM, and survival analysis. This course will make a balance between practical real data analysis with examples and a deeper understanding of the models with mathematical derivations.
Instructor(s): Staff Terms Offered: Winter
Prerequisite(s): STAT 343 (or a similar-level linear regression course) or consent of instructor; comfortable with programming in R.

STAT 34800. Modern Methods in Applied Statistics. 100 Units.
This course covers latent variable models and graphical models; definitions and conditional independence properties; Markov chains, HMMs, mixture models, PCA, factor analysis, and hierarchical Bayes models; methods for estimation and probability computations (EM, variational EM, MCMC, particle filtering, and Kalman Filter); undirected graphs, Markov Random Fields, and decomposable graphs; message passing algorithms; sparse regression, Lasso, and Bayesian regression; and classification generative vs. discriminative. Applications will typically involve high-dimensional data sets, and algorithmic coding will be emphasized.
Terms Offered: Spring

STAT 34900. Data Analysis Project. 100 Units.
The first part of this class will focus on general principles of data analysis and how to report the results of an analysis, including taking account of the context of the data, making informative and clear visual displays, developing relevant statistical models and describing them clearly, and carrying out diagnostic procedures to assess the appropriateness of adopted models. The second half of the class will focus on data analysis projects.
Students working on a data analysis project in another context (e.g., for an MS paper or for consulting) may, with proper permission, use that project for this course as well.
Terms Offered: To be determined
Prerequisite(s): STAT 34700 and permission of instructor

STAT 35201. Introduction to Clinical Trials. 100 Units.
This course will review major components of clinical trial conduct, including the formulation of clinical hypotheses and study endpoints, trial design, development of the research protocol, trial progress monitoring, analysis, and the summary and reporting of results. Other aspects of clinical trials to be discussed include ethical and regulatory issues in human subjects research, data quality control, meta-analytic overviews and consensus in treatment strategy resulting from clinical trials, and the broader impact of clinical trials on public health.
Instructor(s): M. Polley Terms Offered: Winter
Prerequisite(s): PBHS 32100 or STAT 22000; Introductory Statistics or Consent of Instructor
Equivalent Course(s): PBHS 32901

STAT 35410. Genomic Evolution I. 100 Units.
Canalization, a unifying biological principle first enunciated by Conrad Waddington in 1942, is an idea that has had tremendous intellectual influence on developmental biology, evolutionary biology, and mathematics. In this course we will explore canalization in all three contexts through extensive reading and discussion of both the classic and modern primary literature. We intend this exploration to raise new research problems which can be evaluated for further understanding. We encourage participants to present new ideas in this area for comment and discussion.
Instructor(s): M. Long, J. Reinitz, and C-I. Wu Terms Offered: TBD. not offered in 2018-19
Equivalent Course(s): ECEV 35901, EVOL 35901

STAT 35420. Stochastic Processes in Gene Regulation. 100 Units.
This didactic course covers the fundamentals of stochastic chemical processes as they arise in the study of gene regulation. The central object of study is the Chemical Master Equation and its coarse-grainings at the Langevin/ Fokker-Planck, linear noise, and deterministic levels. We will consider both mathematical and computational approaches in contexts where there are both single and multiple deterministic limits.
Instructor(s): J. Reinitz Terms Offered: To be determined
Prerequisite(s): Consent of instructor.
STAT 35450. Fundamentals of Computational Biology: Models and Inference. 100 Units.
Covers key principles in probability and statistics that are used to model and understand biological data. There will be a strong emphasis on stochastic processes and inference in complex hierarchical statistical models. Topics will vary but the typical content would include: Likelihood-based and Bayesian inference, Poisson processes, Markov models, Hidden Markov models, Gaussian Processes, Brownian motion, Birth-death processes, the Coalescent, Graphical models, Markov processes on trees and graphs, Markov Chain Monte Carlo.
Instructor(s): J. Novembre, M. Stephens Terms Offered: Winter
Prerequisite(s): STAT 244
Equivalent Course(s): HGEN 48600

STAT 35460. Fundamentals of Computational Biology: Algorithms and Applications. 100 Units.
This course will cover principles of data structure and algorithms, with emphasis on algorithms that have broad applications in computational biology. The specific topics may include dynamic programming, algorithms for graphs, numerical optimization, finite-difference, schemes, matrix operations/factor analysis, and data management (e.g. SQL, HDF5). We will also discuss some applications of these algorithms (as well as commonly used statistical techniques) in genomics and systems biology, including genome assembly, variant calling, transcriptome inference, and so on.
Instructor(s): Xin He, Mengjie Chen Terms Offered: Spring
Equivalent Course(s): HGEN 48800

STAT 35490. Introduction to Statistical Genetics. 100 Units.
As a result of technological advances over the past few decades, there is a tremendous wealth of genetic data currently being collected. These data have the potential to shed light on the genetic factors influencing traits and diseases, as well as on questions of ancestry and population history. The aim of this course is to develop a thorough understanding of probabilistic models and statistical theory and methods underlying the analysis of genetic data, focusing on problems in complex trait mapping, with some coverage of population genetics. Although the case studies are all in the area of statistical genetics, the statistical inference topics, which will include likelihood-based inference, linear mixed models, and restricted maximum likelihood, among others, are widely applicable to other areas. No biological background is needed, but a strong foundation in linear algebra, as well as probability and statistics at the level of STAT 24400-STAT 24500 or higher is assumed.
Instructor(s): M. McPeek Terms Offered: Autumn
Prerequisite(s): Either HGEN 47100 or both STAT 24400 and 24500. Students without these prerequisites may enroll on a P/NP basis with consent of the instructor.
Equivalent Course(s): STAT 26300

STAT 35500. Statistical Genetics. 100 Units.
This is an advanced course in statistical genetics. We will take an in-depth look at statistical methods development in recent genetics literature, with the aim of achieving a deep understanding of the modeling approaches and assumptions, statistical principles, mathematical theorems, computational issues, and data analytic approaches underlying the methods. The goal is for the student to be able to ultimately apply the principles learned to future statistical methods development for genetic data analysis. This is a discussion course and student presentations will be required. Topics depend on the interests of the participants and will be based on recent published literature. Topics may include, but are not limited to, statistical problems in genetic association mapping, population genetics, integration of different types of genetic data, and genetic models for complex traits. The course material changes every year, and the course may be repeated for credit.
Terms Offered: To be determined
Prerequisite(s): Either HGEN 47100 or both STAT 24400 and 24500. Students without these prerequisites may enroll on a P/NP basis with consent of the instructor.

STAT 35700. Epidemiologic Methods. 100 Units.
This course expands on the material presented in "Principles of Epidemiology," further exploring issues in the conduct of epidemiologic studies. The student will learn the application of both stratified and multivariate methods to the analysis of epidemiologic data. The final project will be to write the "specific aims" and "methods" sections of a research proposal on a topic of the student's choice.
Instructor(s): B. Chiu Terms Offered: Winter
Prerequisite(s): PBHS 30910 AND PBHS 32400 or applied statistics courses through multivariate regression.
Equivalent Course(s): PBHS 31001

STAT 35800. Statistical Applications. 100 Units.
This course provides a transition between statistical theory and practice. The course will cover statistical applications in medicine, mental health, environmental science, analytical chemistry, and public policy. Lectures are oriented around specific examples from a variety of content areas. Opportunities for the class to work on interesting applied problems presented by U of C faculty will be provided. Although an overview of relevant statistical theory will be presented, emphasis is on the development of statistical solutions to interesting applied problems.
Instructor(s): R. Gibbons Terms Offered: Autumn
Prerequisite(s): PBHS 32700/STAT 22700 or STAT 34700 or consent of instructor.
Equivalent Course(s): CHDV 32702, PBHS 33500
STAT 35920. Applied Bayesian Modeling and Inference. 100 Units.
Course begins with basic probability and distribution theory, and covers a wide range of topics related to Bayesian modeling, computation, and inference. Significant amount of effort will be directed to teaching students on how to build and apply hierarchical models and perform posterior inference. The first half of the course will be focused on basic theory, modeling, and computation using Markov chain Monte Carlo methods, and the second half of the course will be about advanced models and applications. Computation and application will be emphasized so that students will be able to solve real-world problems with Bayesian techniques.
Instructor(s): Y. Ji Terms Offered: Winter
Prerequisite(s): STAT 24400 and STAT 24500 or master level training in statistics.
Equivalent Course(s): PBHS 43010

STAT 36600. Decision Theory. 100 Units.
This course covers statistical decision theory with examples drawn from modern high-dimensional and nonparametric estimation. Topics that will be covered include basic information theory, decision theory, asymptotic equivalence, Gaussian sequence model, sparse regression, model selection, aggregation, and large covariance matrix estimation. Lower bound techniques such as Bayes, Le Cam, and Fano's methods will be taught.
Terms Offered: To be determined

STAT 36700. History of Statistics. 100 Units.
This course covers topics in the history of statistics, from the eleventh century to the middle of the twentieth century. We focus on the period from 1650 to 1950, with an emphasis on the mathematical developments in the theory of probability and how they came to be used in the sciences. Our goals are both to quantify uncertainty in observational data and to develop a conceptual framework for scientific theories. This course includes broad views of the development of the subject and closer looks at specific people and investigations, including reanalyses of historical data.
Instructor(s): S. Stigler Terms Offered: Spring
Prerequisite(s): Prior statistics course
Equivalent Course(s): STAT 26700, HIPS 25600, CHSS 32900

STAT 36900. Applied Longitudinal Data Analysis. 100 Units.
Longitudinal data consist of multiple measures over time on a sample of individuals. This type of data occurs extensively in both observational and experimental biomedical and public health studies, as well as in studies in sociology and applied economics. This course will provide an introduction to the principles and methods for the analysis of longitudinal data. Whereas some supporting statistical theory will be given, emphasis will be on data analysis and interpretation of models for longitudinal data. Problems will be motivated by applications in epidemiology, clinical medicine, health services research, and disease natural history studies.
Instructor(s): D. Hedeker Terms Offered: Spring
Prerequisite(s): PBHS 32400/STAT 22400 or equivalent, and PBHS 32600/STAT 22600 or PBHS 32700/STAT 22700 or equivalent; or consent of instructor.
Equivalent Course(s): CHDV 32501, PBHS 33300

STAT 37400. Nonparametric Inference. 100 Units.
Nonparametric inference is about developing statistical methods and models that make weak assumptions. A typical nonparametric approach estimates a nonlinear function from an infinite dimensional space rather than a linear model from a finite dimensional space. This course gives an introduction to nonparametric inference, with a focus on density estimation, regression, confidence sets, orthogonal functions, random processes, and kernels. The course treats nonparametric methodology and its use, together with theory that explains the statistical properties of the methods.
Instructor(s): Staff Terms Offered: Autumn
Prerequisite(s): STAT 24400 or STAT 24410 w/B- or better is required; alternatively STAT 22400 w/B+ or better and exposure to multivariate calculus (MATH 16300 or MATH 16310 or MATH 18400 or MATH 19520 or MATH 20000 or MATH 20500 or MATH 20510 or MATH 20800) and linear algebra (MATH 19620 or MATH 20250 or STAT 24300 or equivalent). Master's students in Statistics can enroll without prerequisites.
Equivalent Course(s): STAT 27400

STAT 37411. Topological Data Analysis. 100 Units.
Topological data analysis seeks to understand and exploit topology when exploring and learning from data. This course surveys core ideas and recent developments in the field and will prepare students to use topology in data analysis tasks. The core of the course will include computation with topological spaces, the mapper algorithm, and persistent homology, and cover theoretical results, algorithms, and a variety of applications. Additional topics from algebraic topology, metric geometry, category theory, and quiver representation theory will be developed from applied and computational perspectives.
Terms Offered: Winter
Prerequisite(s): Linear algebra, prior programming experience, exposure to graph theory/algorithms.
Equivalent Course(s): CAAM 37411

STAT 37601. Machine Learning and Large-Scale Data Analysis. 100 Units.
This course is an introduction to machine learning and the analysis of large data sets using distributed computation and storage infrastructure. Basic machine learning methodology and relevant statistical theory
will be presented in lectures. Homework exercises will give students hands-on experience with the methods on different types of data. Methods include algorithms for clustering, binary classification, and hierarchical Bayesian modeling. Data types include images, archives of scientific articles, online ad clickthrough logs, and public records of the City of Chicago. Programming will be based on Python and R, but previous exposure to these languages is not assumed.

Instructor(s): Staff
Prerequisite(s): CMSC 15400 or CMSC 12200 and STAT 22000 or STAT 23400, or by consent.
Note(s): The prerequisites are under review and may change.
Equivalent Course(s): CMSC 25025

STAT 37710. Machine Learning. 100 Units.
This course provides hands-on experience with a range of contemporary machine learning algorithms, as well as an introduction to the theoretical aspects of the subject. Topics covered include: the PAC framework, Bayesian learning, graphical models, clustering, dimensionality reduction, kernel methods including SVMs, matrix completion, neural networks, and an introduction to statistical learning theory.
Terms Offered: To be determined
Prerequisite(s): Consent of instructor
Equivalent Course(s): CMSC 35400, CAAM 37710

STAT 37790. Topics in Statistical Machine Learning. 100 Units.
Topics in Statistical Machine Learning" is a second graduate level course in machine learning, assuming students have had previous exposure to machine learning and statistical theory. The emphasis of the course is on statistical methodology, learning theory, and algorithms for large-scale, high dimensional data. The selection of topics is influenced by recent research results, and students can take the course in more than one quarter.
Terms Offered: To be determined
Equivalent Course(s): CMSC 35425

STAT 37791. Topics in Machine Learning. 100 Units.
This course covers selected topics in dimension reduction, randomized algorithm, sparsity, convex optimization, and deep learning, with a focus on scientific computing.
Terms Offered: Spring
Prerequisite(s): Enrolled PhD or MS student in Statistics or in Computational and Applied Mathematics, or consent of instructor.
Note(s): Recommended prerequisites: STAT 30900, STAT 31015, and undergraduate probability.
Equivalent Course(s): CAAM 37791

STAT 37792. Topics in Deep Learning: Generative Models. 100 Units.
This course will be a hands on exploration of various approaches to generative modeling with deep networks. Topics include variational auto encoders, flow models, GAN models, and energy models. Participation in this course requires familiarity with pytorch and a strong background in statistical modeling. The course will primarily consist of paper presentations. Each presenter would be required to report on experiments performed with the algorithm proposed in the paper, exploring strengths and weaknesses of the methods.
Instructor(s): Y. Amit Terms Offered: Autumn
Prerequisite(s): STAT 34300, STAT 34700, STAT 34800, and STAT 37601/CMSC 25025, or STAT 37710/CMSC 35400

STAT 37793. Topics in Deep Learning: Discriminative Models. 100 Units.
This course will explore modern approaches to optimization, data augmentation, and domain shift for deep neural networks from both theoretical and empirical perspectives. Participation will require independent investigation with PyTorch as well as paper presentations.
Terms Offered: Winter
Prerequisite(s): STAT 37601 or STAT 37710 or consent of instructor.

STAT 37794. Special Topics in Machine Learning. 100 Units.
Learned emulators leverage neural networks to increase the speed of physics simulations in climate models, astrophysics, high-energy physics, and more. Recent empirical results have illustrated that these emulators can speed up traditional simulations by up to eight orders of magnitude. However, little is understood about these emulators. While it is possible that recent results are representative of what is possible in most settings, a more likely scenario is that these approaches are more effective for some simulators than others, and that learned emulators achieve strong average-case performance but fail to capture rare but important phenomena. In this graduate seminar course we will provide an overview and investigate recent literature on this topic, focusing on the following questions: 1. Introduction to learned emulators: how do they work, where have they been successful so far and what are the goals in this field? 2. Two different paradigms of learned emulation: physics vs. data driven. What are the advantages and pitfalls of each? 3. Robustness of emulation to noise: what is known so far? 4. Parameter estimation: how to handle parameter uncertainty? We will provide a list of papers covering the above topics and students will be evaluated on in-class presentations.
Instructor(s): Dana Mendelson (Math) and Rebecca Willett (CS/Stats) Terms Offered: Autumn
Prerequisite(s): Students should be familiar with a numerical programming language like Python, Julia, R, or Matlab and the content of CMSC 35400. Students should also have familiarity with the contents of MATH 27300 and MATH 27500 or similar.
Note(s): Because this is a seminar course, it will be capped at 15 students, 4 Math, 4 CS/Stats, and 7 with instructor permission.
Equivalent Course(s): CAAM 37794, CMSC 35490, MATH 37794

**STAT 37795. Causal Inference with Machine Learning. 100 Units.**
This course is an introduction to causal inference from observational data with an emphasis on the role of machine learning. We'll cover both foundations of causality—causal graphs and identification results—and practical aspects of causal estimation. Time permitting, we'll also look at how causal formalism affects the practice of machine learning.
Instructor(s): V. Veitch Terms Offered: Spring
Prerequisite(s): Background in machine learning

**STAT 37796. Topics in Machine Learning: Symmetries and Harmonic Analysis. 100 Units.**
TBD
Instructor(s): R. Kondor Terms Offered: Autumn

**STAT 37797. Topics in Mathematical Data Science: Spectral Methods and Nonconvex Optimization. 100 Units.**
TBD
Terms Offered: Autumn

**STAT 37810. Statistical Computing A. 50 Units.**
Statistical Computing A focuses on common data technology used in statistical computing and broader data science. The course takes place in the second half of the autumn quarter, after STAT 37810 (Statistical Computing A). Topics include storage and accessing of large data; basic working knowledge of relational database and its querying language SQL; introduction to distributed file system and example usage of Hadoop; Python and its applications in text analysis; access and usage of high-performance computer clusters, rudimentary parallel computing, web data access. XML and Javascript may be used occasionally. A short introduction to SAS will be given if time permits. The main computing software will be Python with some R.
Terms Offered: Autumn
Prerequisite(s): Instructor consent. STAT 37810 recommended.

**STAT 37820. Statistical Computing B. 50 Units.**
This course is an introduction to statistical programming in R. Students will learn how to design, write, debug and test functions by implementing several famous algorithms in statistics such as Gibbs Sampling and Expectation Maximization. A basic familiarity with R is needed, but no prior programming experience is required. The course will also introduce students to the use of version control with Git and consider the differences and similarities between R and Python.
Terms Offered: Autumn
Prerequisite(s): Instructor consent.

**STAT 37830. Scientific Computing with Python. 100 Units.**
This course provides a detailed, rigorous treatment of probability from the point of view of measure theory, as well as existence theorems, integration and expected values, characteristic functions, moment problems, limit laws, Radon-Nikodym derivatives, and conditional probabilities.
Terms Offered: Winter
Prerequisite(s): STAT 30400 or consent of instructor

**STAT 38100. Measure-Theoretic Probability I. 100 Units.**
This course provides a detailed, rigorous treatment of probability from the point of view of measure theory, as well as existence theorems, integration and expected values, characteristic functions, moment problems, limit laws, Radon-Nikodym derivatives, and conditional probabilities.
Terms Offered: Winter
Prerequisite(s): STAT 30400 or consent of instructor

**STAT 38300. Measure-Theoretic Probability III. 100 Units.**
This course continues material covered in STAT 38100, with topics that include Lp spaces, Radon-Nikodym theorem, conditional expectation, and martingale theory.
Terms Offered: Spring
Prerequisite(s): STAT 38100

**STAT 38510. Brownian Motion and Stochastic Calculus. 100 Units.**
This is a rigorous introduction to the mathematical theory of Brownian motion and the corresponding integration theory (stochastic integration). This is material that all analysis graduate students should learn at some point whether or not they are immediately planning to use probabilistic techniques. It is also a natural course for more advanced math students who want to broaden their mathematical education and to increase their marketability for nonacademic positions. In particular, it is one of the most fundamental mathematical tools used in financial
mathematics (although we will not discuss finance in this course). This course differs from the more applied
STAT 39000 in that concepts are developed precisely and rigorously.
Terms Offered: To be determined.
Note(s): Recommended prerequisites: STAT 38300; or MATH 31200, MATH 31300, and MATH 31400; or consent
of instructor.
Equivalent Course(s): MATH 38511

STAT 38520. Topics in Random Matrix Theory. 100 Units.
Random matrix theory (RMT) is among the most prominent subjects in modern probability theory, with
applications in a wide range of disciplines (including physics, statistics, engineering, and finance). The purpose
of this course is to study a broad sample of the most prominent research programs in RMT as well as their
motivating applications. Main topics will include (time permitting) the moment method in RMT and its
connection to combinatorics, universality, operator limits, and matrix concentration.
Terms Offered: Winter
Prerequisite(s): PhD student in Statistics or Math or Computational and Applied Mathematics or TTIC or MS
student in Statistics or Computational and Applied Mathematics. Other students may enroll with consent of
instructor.
Note(s): Prerequisite notes: Graduate or advanced undergraduate probability theory and undergraduate linear
algebra and combinatorics are recommended.
Equivalent Course(s): CAAM 38520

STAT 39000. Stochastic Calculus. 100 Units.
The course starts with a quick introduction to martingales in discrete time, and then Brownian motion and the
Ito integral are defined carefully. The main tools of stochastic calculus (Ito’s formula, Feynman-Kac formula,
Girsanov theorem, etc.) are developed. The treatment includes discussions of simulation and the relationship
with partial differential equations. Some applications are given to option pricing, but much more on this is done
in other courses. The course ends with an introduction to jump process (Levy processes) and the corresponding
integration theory. Program requirement.
Instructor(s): G. Lawler Terms Offered: Winter
Equivalent Course(s): FINM 34500

STAT 39010. Stochastic Calculus I. 50 Units.
The course starts with a quick introduction to martingales in discrete time, and then Brownian motion and the
Ito integral are defined carefully. The main tools of stochastic calculus (Ito’s formula, Feynman-Kac formula,
Girsanov theorem, etc.) are developed. The treatment includes discussions of simulation and the relationship
with partial differential equations. Some applications are given to option pricing, but much more on this is done
in other courses. The course ends with an introduction to jump process (Levy processes) and the corresponding
integration theory.
Terms Offered: Winter
Prerequisite(s): Consent of instructor.
Equivalent Course(s): FINM 34510

STAT 39020. Stochastic Calculus II. 50 Units.
Continuation of FINM 34510
Instructor(s): Prof. Greg Lawler Terms Offered: Winter
Prerequisite(s): FINM 34510
Equivalent Course(s): FINM 34520

STAT 39800. Field Research. 300.00 Units.
This Summer Quarter course offers graduate students in the Statistics Department the opportunity to apply
statistics knowledge that they have acquired to a real industry or business situation. During the summer quarter
in which they are registered for the course, students complete a paid or unpaid internship of at least six weeks.
Prior to the start of the work experience, students secure faculty consent for an independent study project to be
completed during the internship quarter.
Terms Offered: Summer only
Prerequisite(s): Masters or PhD student in Statistics or Consent of instructor and faculty advisor.

STAT 39900. Masters Seminar: Statistics. 300.00 Units.
This course is for Statistics Master’s students to carry out directed reading or guided work on topics related to
their Master’s papers.
Prerequisite(s): Masters or PhD student in Statistics

STAT 40100. Reading/Research: Statistics. 300.00 Units.
This course allows doctoral students to receive credit for advanced work related to their dissertation topics.
Students register for one of the listed faculty sections with prior consent from the respective instructor. Students
may work with faculty from other departments; however, they still must obtain permission from and register
with one of the listed faculty members in the Department of Statistics.
Terms Offered: All quarters
Prerequisite(s): Masters or PhD student in Statistics or consent of instructor
STAT 41500-41600. High-Dimensional Statistics I-II.
These courses treat statistical problems where the number of variables is very large. Classical statistical methods and theory often fail in such settings. Modern research has begun to develop techniques that can be effective in high dimensions, and that can be understood theoretically. The first quarter introduces a range of statistical frameworks for finding low-dimensional structure in high-dimensional data, such as sparsity in regression, sparse graphical models, or low-rank structure. This quarter emphasizes methods for estimation and inference developed in these areas, along with theoretical analysis of their properties. The second quarter emphasizes foundational aspects of high-dimensional statistics, focusing on principles that are used across a range of problems and are likely to be relevant for methods developed in the future. Topics include "the curse of dimensionality," elements of random matrix theory, properties of high-dimensional covariance matrices, concentration of measure, dimensionality reduction techniques, and handling miss-specified models. The courses may be taken separately.

STAT 41500. High-Dimensional Statistics I. 100 Units.
These courses treat statistical problems where the number of variables is very large. Classical statistical methods and theory often fail in such settings. Modern research has begun to develop techniques that can be effective in high dimensions, and that can be understood theoretically. The first quarter introduces a range of statistical frameworks for finding low-dimensional structure in high-dimensional data, such as sparsity in regression, sparse graphical models, or low-rank structure. This quarter emphasizes methods for estimation and inference developed in these areas, along with theoretical analysis of their properties. The second quarter emphasizes foundational aspects of high-dimensional statistics, focusing on principles that are used across a range of problems and are likely to be relevant for methods developed in the future. Topics include "the curse of dimensionality," elements of random matrix theory, properties of high-dimensional covariance matrices, concentration of measure, dimensionality reduction techniques, and handling miss-specified models. The courses may be taken separately.

Terms Offered: To be determined
Prerequisite(s): STAT 30100 and STAT 30400 and STAT 31015, or consent of instructor

STAT 41600. High-Dimensional Statistics II. 100 Units.
These courses treat statistical problems where the number of variables is very large. Classical statistical methods and theory often fail in such settings. Modern research has begun to develop techniques that can be effective in high dimensions, and that can be understood theoretically. The first quarter introduces a range of statistical frameworks for finding low-dimensional structure in high-dimensional data, such as sparsity in regression, sparse graphical models, or low-rank structure. This quarter emphasizes methods for estimation and inference developed in these areas, along with theoretical analysis of their properties. The second quarter emphasizes foundational aspects of high-dimensional statistics, focusing on principles that are used across a range of problems and are likely to be relevant for methods developed in the future. Topics include "the curse of dimensionality," elements of random matrix theory, properties of high-dimensional covariance matrices, concentration of measure, dimensionality reduction techniques, and handling miss-specified models. The courses may be taken separately.

Terms Offered: To be determined
Prerequisite(s): STAT 30100 or STAT 30400 or STAT 31015, or consent of instructor

STAT 41510. Bayesian Nonparametrics. 100 Units.
Bayesian nonparametric methods are increasingly important tools in machine learning and statistics. We will discuss nonparametric Bayesian approaches to mixture models, latent feature models, hierarchical models, network models, and high-dimensional regression models. Topics that will be covered include Dirichlet process, Chinese restaurant process, Pitman-Yor process, Indian buffet process, Gaussian process, and their computational techniques via Gibbs sampling and variational inference. Frequentist evaluations of posterior distributions will also be discussed in nonparametric and high-dimensional settings.

Instructor(s): C. Gao
Terms Offered: To be determined
Prerequisite(s): STAT 30200

STAT 41511. Topics in Robust and Semiparametric Statistics. 100 Units.
This course is about statistical estimation and inference with nuisance parameters. Examples include location estimation with unknown density, Cox proportional hazard model, low-dimensional inference in sparse regression, and robust estimation with arbitrary contamination. We will learn tangent spaces, efficient score functions, and information operators. Basic empirical process tools will also be discussed.

Terms Offered: Spring
Prerequisite(s): STAT 30100

STAT 41520. Topics in Selective Inference. 100 Units.
This course will study the problem of selective inference where we would like to provide statistical guarantees about hypotheses or parameters whose definitions are influenced by our analysis of the same data set. Performing valid inference is challenging since we must find a way to condition on the outcome of the selection process which is not always simple to characterize. The course will discuss both recent advances and open problems in this field.

Instructor(s): R. Barber
Terms Offered: To be determined
Prerequisite(s): STAT 27850/30850 or STAT 30200 or consent of instructor
STAT 41530. **Topics in Causal Inference. 100 Units.**
We will start with a light and comparative introduction of two causal inference languages: the potential outcome model and the graphical representation of causal effects. In the course, we will discuss topics including confounding, instrumental variables (IV), mediation analysis, and effective treatment allocations, with their applications in genetics and epidemiological research.

STAT 42600. **Theoretical Neuroscience: Statistics and Information Theory. 100 Units.**
This course begins with an introduction to inference and statistical methods in data analysis. We then cover the two main sections of the course: I) Encoding and II) Decoding in single neurons and neural populations. The encoding section will cover receptive field analysis (STA, STC and non-linear methods such as maximally informative dimensions) and will explore linear-nonlinear-Poisson models of neural encoding as well as generalized linear models alongside newer population coding models. The decoding section will cover basic methods for inferring stimuli or behaviors from spike train data, including both linear and correlational approaches to population decoding. The course will use examples from real data (where appropriate) in the problem sets which students will solve using MATLAB.
Instructor(s): S. Palmer Terms Offered: Spring
Prerequisite(s): Prior exposure to basic calculus and probability theory, CPNS 35500 or instructor consent.
Equivalent Course(s): CPNS 35600, ORGB 42600

STAT 44100. **Consulting In Statistics. 300.00 Units.**
This seminar course is an internal training program for graduate students in Statistics. The primary goal is to expose the students to applications that involve statistical thinking and to have hands on experience on real world data. The projects are provided by researchers from the university community. Participating students form teams to work on selected projects under faculty guidance and to present their work to all student consultants and researcher clients.

STAT 70000. **Advanced Study: Statistics. 300.00 Units.**
Advanced Study: Statistics