

# Department of Statistics

---

## Chair

- Dan Liviu Nicolae, Statistics and Medicine

## Professors

- Yali Amit
- Mihai Anitescu, Argonne National Laboratory
- Guillaume Bal
- Lars Peter Hansen, Economics and Statistics
- Steven P. Lalley
- Gregory F. Lawler, Mathematics and Statistics
- Peter McCullagh
- Mary Sara McPeck
- Per Mykland
- Dan Liviu Nicolae, Statistics and Medicine
- John Reinitz
- Mary Silber
- Michael L. Stein, Master of the Physical Sciences Collegiate Division
- Matthew Stephens
- Stephen M. Stigler
- Ronald Thisted, Vice Provost for Academic Affairs, Public Health Sciences, Statistics
- Kirk M. Wolter
- Wei Biao Wu

## Associate Professors

- Lek-Heng Lim
- Jonathan Weare

## Assistant Professors

- Rina Foygel Barber
- Chao Gao
- Zheng (Tracy) Ke
- Imre Risi Kondor, Computer Science and Statistics

## Senior Lecturers

- Linda Brant Collins
- Mei Wang

## Lecturers

- Kendra S. Burbank
- Yibi Huang

## Visiting Professors

- James O. Berger

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.

#### **Programs and Requirements for the Ph.D.**

All sufficiently well-prepared students take 3 of 4 sequences in their first year:

- Applied Statistics
- Theoretical Statistics
- Probability
- Computation and Machine Learning

All students pass prelim exams in 2 of the 4 subjects by the beginning of their second year. Well-prepared students may be allowed to pass one or both of their exams upon arrival. Students should take a distribution requirement of up to two courses in their second year and are otherwise encouraged to explore the great variety of graduate courses on offer, both inside the department and in other departments.

Starting in their second year, students should find a topic for a Ph.D. dissertation and establish a relationship with a Ph.D. adviser. Taking courses with potential advisers is part of this process. The detailed process is listed here ([http://www.stat.uchicago.edu/students/phd\\_rules.shtml](http://www.stat.uchicago.edu/students/phd_rules.shtml)).

#### **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 (<http://www.stat.uchicago.edu/seminars/index.shtml>).

Ph.D. students should also participate in the department's consulting program (<http://galton.uchicago.edu/consulting/index.shtml>), 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 (<http://galton.uchicago.edu/students/master.shtml>). (<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 (<http://galton.uchicago.edu/seminars/index.shtml>), computation (<http://galton.uchicago.edu/local/computing>), and student and faculty offices (<http://galton.uchicago.edu/people>)—are located in Eckhart Hall or neighboring Ryerson Hall. Each student is assigned a desk in one of several offices. A small departmental library and conference room is a common meeting place for formal and informal gatherings of students and faculty. 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 (<http://www.stat.uchicago.edu/local/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 Health Studies (<http://health.bsd.uchicago.edu>), Human Genetics (<http://genes.uchicago.edu>), Financial Mathematics and Econometrics (<http://stevanovichcenter.uchicago.edu/page/seminars>), 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 (<http://galton.uchicago.edu/seminars/index.shtml>) 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. Concurrent or prior linear algebra (MATH 19620 or 20250 or STAT 24300 or equivalent) is recommended for students continuing to STAT 24510.

Note(s): Some previous experience with statistics helpful but not required. 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 30040. Statistical Theory and Methods IIa. 100 Units.

This course is a continuation of STAT 24410. The focus is on theory and practice of linear models, including the analysis of variance, regression, correlation, and some multivariate analysis. Additional topics may include bootstrapping for regression models, nonparametric regression, and regression models with correlated errors.

Terms Offered: May be offered in Winter.

Prerequisite(s): STAT 24410. Linear algebra (MATH 19620 or 20250 or STAT 24300 or equivalent).

Note(s): Students may count either STAT 24500 or STAT 24510, but not both, toward the forty-two credits required for graduation.

Equivalent Course(s): STAT 24510

#### STAT 30100. Mathematical Statistics I. 100 Units.

This course is part of a two-quarter sequence on the theory of statistics. Topics will include exponential, curved exponential, and location-scale families; mixtures, hierarchical and conditional modeling including compatibility of conditional distributions; principles of estimation; identifiability, sufficiency, minimal sufficiency, ancillarity, completeness; properties of the likelihood function and likelihood-based inference, both univariate and multivariate, including examples in which the usual regularity conditions do not hold; elements of Bayesian inference and comparison with frequentist methods; and multivariate information inequality. Part of the course will be devoted to elementary asymptotic methods that are useful in the practice of statistics, including methods to derive asymptotic distributions of various estimators and test statistics, such as Pearson's chi-square, standard and nonstandard asymptotics of maximum likelihood estimators and Bayesian estimators, asymptotics of order statistics and extreme order statistics, Cramer's theorem including situations in which the second-order term is needed, and asymptotic efficiency. Other topics (e.g., methods for dependent observations) may be covered if time permits.

Instructor(s): Staff Terms Offered: Winter

Prerequisite(s): STAT 30400 or consent of instructor

#### STAT 30200. Mathematical Statistics II. 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: Spring

Prerequisite(s): STAT 30100 and STAT 30400, or consent of instructor

**STAT 30210. Bayesian Analysis and Principles of Statistics. 100 Units.**

This course continues the development of Mathematical Statistics, with an emphasis on Bayesian analysis and underlying principles of inference. Topics include Bayesian Inference and Computation, Frequentist Inference and interpretation of  $p$  values and confidence intervals, Decision theory, admissibility and Stein's paradox, the Likelihood principle, Exchangeability and De Finetti's theorem, hierarchical modelling, multiple comparisons and False Discovery Rates. The mathematical level will generally be at that of an easy advanced calculus course. We will assume familiarity with standard statistical distributions (e.g., Normal, Poisson, Binomial, Exponential), with the laws of probability, expectation, conditional expectation, etc., and exposure to common statistical concepts such as  $p$  values and confidence intervals. Familiarity with the R statistical language will also be expected, and homework assignments will include programming problems in R.

Terms Offered: Spring

Prerequisite(s): STAT 30400 or consent of instructor

**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 and MATH 20500, or consent of instructor

**STAT 30600. Advanced Statistical Inference I. 100 Units.**

Topics covered in this course will include: Gaussian distributions: conditional distributions; maximum likelihood and REML; Laplace approximation and associated expansion; combinatorics and the partition lattice; Mobius inversion; moments, cumulants symmetric functions and  $S_k$ -statistics; cluster expansions; Bartlett identities and Bartlett adjustment; random partitions, partition processes, CRP process; Gauss-Ewens cluster process: classification models; trees rooted and unrooted; exchangeable random trees; Cox processes used for classification.

Terms Offered: Autumn, Spring, Autumn or Spring

Prerequisite(s): Consent of instructor

Note(s): May not be offered in 2016-17

**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 19520 or 20000 or 20500 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: Spring

Prerequisite(s): STAT 30400, 30100, and 30210, or consent of instructor

**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.

Instructor(s): R. Barber Terms Offered: Winter

Prerequisite(s): Stat 24400 or equivalent.

Equivalent Course(s): STAT 27850

**STAT 3090. 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): CMSC 37810

**STAT 31015. Mathematical Computation IIA: Convex Optimization. 100 Units.**

This course covers the fundamentals of convex optimization. Topics will include basic convex geometry and convex analysis, KKT condition, Fenchel and Lagrange duality theory; six standard convex optimization problems and their properties and applications: linear programming, geometric programming, second-order cone programming, semidefinite programming, linearly and quadratically constrained quadratic programming. In the last part of the course we will examine the generalized moment problem --- a powerful technique that allows one to encode a wide variety of problems (in probability, statistics, control theory, financial mathematics, signal processing, etc) and solve them or their relaxations as convex optimization problems.

Terms Offered: Winter

Prerequisite(s): STAT 30900/CMSC 37810

**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

Note(s): Not offered in 2016-17, expected to be offered in 2017-18.

**STAT 31061. Further Mathematical Computation: Matrix Computation. 100 Units.**

This course is primarily about iterative algorithms in matrix computation. For linear systems and least squares problems, we will discuss stationary methods (Jacobi, Gauss-Seidel, SOR), semi-iterative methods (Richardson, steepest descent, Chebyshev, conjugate gradient), and Krylov subspace methods (MINRES, SYMMLQ, LSQR, GMRES, QMR, BiCG). We will cover some basic ideas for preconditioning and stopping conditions. For eigenvalue problems, we will discuss direct (Givens and Householder) and iterative (Lanczos and Arnoldi) methods for reducing a matrix into tridiagonal and Hessenberg forms, as well as power, inverse power, Rayleigh quotient, Jacobi, Jacobi-Davidson, and Francis QR algorithms for extraction of eigenvalues/eigenvectors. Lastly, we will discuss algorithms for generalized and quadratic eigenvalue problems (QZ algorithm) as well as for singular value decomposition (Golub-Kahan and Golub-Reinsch).

Terms Offered: Winter

Prerequisite(s): STAT 30900/CMSC 37810

Note(s): Not offered in 2016-17

**STAT 31095. Numeric Solution of Ordinary Differential Equations. 100 Units.**

This course will cover numerical methods for solving ordinary differential equations. Topics will include the development and analysis of Runge-Kutta and multistep methods, methods for stiff problems, and adaptive methods such as embedded Runge-Kutta. Additional topics such as symplectic methods, methods for boundary value problems, and methods for differential algebraic equations may also be covered, depending on the interests of the students. Coursework will include both computation and analysis. Theoretical results will be illustrated by numerical experiments on simple systems from celestial mechanics, molecular dynamics, chemical kinetics, and other fields. No knowledge of differential equations or numerical analysis will be assumed.

Instructor(s): B. Van Koten Terms Offered: Autumn

Prerequisite(s): Linear algebra (MATH 19620 or STAT 24300, or equivalent) and multivariate calculus (MATH 19520 or 20000, or equivalent), or consent of instructor.

Note(s): Not offered in 2016-17

**STAT 31100. Mathematical Computation III: Numerical Methods for PDE's. 100 Units.**

This course covers the major classes of numerical methods used for solving most of the partial differential equations that arise in science and engineering. Topics: Finite differences for elliptic, parabolic, and hyperbolic equations. Iterative methods for linear systems (CG, GMRES). Finite elements. Finite volumes for conservation laws. Spectral methods. Reformulation of PDE as boundary integral equations. Fast algorithms including the fast multipole method. The evaluation will be a mix of theoretical and programming exercises, as well as a project of the student's choice.

Terms Offered: Spring

Prerequisite(s): Numerical linear algebra at the level of STAT 24300/30750, and basic Fourier series.

**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: Autumn

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: Winter

**STAT 31300. Introduction to Stochastic Processes II. 100 Units.**

Topics include continuous-time Markov chains, Markov chain Monte Carlo, discrete-time martingales, and Brownian motion and diffusions. Our emphasis is on defining the processes and calculating or approximating various related probabilities. The measure theoretic aspects of these processes are not covered rigorously.

Terms Offered: Spring

Prerequisite(s): STAT 31200 or consent of instructor

Note(s): Not offered in 2014-15

**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.

Instructor(s): Staff Terms Offered: Autumn

Prerequisite(s): Multivariate calculus and linear algebra

**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.

Instructor(s): Staff Terms Offered: Winter

Prerequisite(s): Multivariate calculus and linear algebra

**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: May be offered in Winter

Prerequisite(s): STAT 24400 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 is a prerequisite. This course is a pre-requisite for "Advanced Topics in Causal Inference" and "Mediation, moderation, and spillover effects."

Instructor(s): K. Yamaguchi Terms Offered: Winter

Prerequisite(s): Intermediate Statistics or equivalent such as STAT 224/PBHS 324, PP 31301, BUS 41100, or SOC 30005 is a prerequisite.

Note(s): Graduate course, open to advanced undergraduates. CHDV Distribution: M, M\*

Equivalent Course(s): SOCI 30315, PBHS 43201, PLS 30102, 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](http://boothportal.chicagobooth.edu/portal/server.pt/community/course_search)

Terms Offered: Autumn

Prerequisite(s): One year of calculus

Equivalent Course(s): BUSF 41901

**STAT 32500. Statistical Inference. 100 Units.**

This Ph.D.-level course is the second in a two-quarter sequence with Business 41901/Statistics 32400. The central topic is statistical inference. The course will focus on inference issues in a variety of linear models. The key models that will be covered are the linear regression model, linear panel data models, and the linear instrumental variable model. The focus of the course will be on developing tools for performing classical inference within these models. We will cover basic asymptotic theory, estimation of covariance matrices allowing for heteroskedasticity and dependence, and the bootstrap. The basics of generalized method of moments will be covered in the context of the linear instrumental variables model. There will also be some discussion of Bayesian inference and finite-sample classical inference.

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](http://boothportal.chicagobooth.edu/portal/server.pt/community/course_search)

Terms Offered: Winter

Prerequisite(s): BUSF 41901/STAT 32400

Equivalent Course(s): BUSF 41902

**STAT 32600. Marketing Topics: Bayesian Applications in Marketing and Micro Econometrics. 100 Units.**

This course covers some key topics at the research frontier in quantitative marketing. We formulate and estimate models of consumer decision-making, and then explore the normative and positive consequences of the inferred consumer behavior for optimal marketing decisions and market structure. Topics include: Foundations of demand modeling, measurement of consumer heterogeneity, the origin and evolution of preferences, state dependence in demand, dynamic discrete choice models, learning and memory models, storable goods demand, diffusion models and durable goods demand, stated choice models, advertising dynamics, and search and shopping behavior. 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](http://boothportal.chicagobooth.edu/portal/server.pt/community/course_search)

Terms Offered: Spring

Equivalent Course(s): BUSF 37904

**STAT 32900. Applied Multivariate Analysis. 100 Units.**

The course will introduce the basic theory and applications for analyzing multi-dimensional 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.

Terms Offered: Spring

Prerequisite(s): STAT 24400-24500 or BUSF 41901/STAT 32400 or BUSF 41902/STAT 32500 or equivalent courses

Equivalent Course(s): BUSF 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, linear discriminant analysis, canonical correlation analysis, cluster analysis, etc. Understanding these techniques require some facility with matrices in addition to some basic statistics, both of which the student will acquire during the course. *Program elective.*

Instructor(s): L. Lim Terms Offered: Autumn

Equivalent Course(s): FINM 33180

**STAT 32950. Multivariate Statistical Analysis: Applications and Techniques. 100 Units.**

This course focuses on applications and techniques for analysis of multivariate and high dimensional data. Beginning subjects cover common multivariate techniques and dimension reduction, including principal component analysis, factor model, canonical correlation, multi-dimensional 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.

Terms Offered: Spring

Prerequisite(s): STAT 24400-24500 or STAT 24410-24510 or consent of instructor

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.

Terms Offered: Autumn

Prerequisite(s): Consent of instructor

**STAT 33500. Time-Series Analysis/Forecast. 100 Units.**

Forecasting plays an important role in business planning and decision-making. 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/](http://boothportal.chicagobooth.edu/portal/server.pt/community/course_search/)

Terms Offered: Winter

Prerequisite(s): BUSF 41901/STAT 32400 or instructor consent

Equivalent Course(s): BUSF 41910

**STAT 33560. Chaos and Predictability. 100 Units.**

This course explores the connection between our models of the world and our observations of it. Theoretical questions of predictability as well as applied methods of forecasting are developed. By adopting a geometric approach to the analysis of dynamical systems, traditional linear analysis of time series is seen be a special case of the more general nonlinear approach. The analysis of time series both from chaotic systems and from nonlinear stochastic systems is used to exemplify the strengths, weaknesses and risks of applying linear intuitions in a nonlinear context. Techniques of forecast evaluation are considered and illustrated with examples from several fields including weather, finance and medicine. The student will develop a software toolkit for the analysis and modelling. Using this toolkit, the efficacy of modern methods for analysis and prediction is considered both in mathematical systems and in real systems. A basic proficiency in a statistical computing (MATLAB, Mathematica, or R, for example) is needed, but no complex programming is required. Undergraduates with a solid background in calculus and one or more classes in statistics are welcome.

Terms Offered: Spring

Prerequisite(s): STAT 24500 or equivalent (can be taken concurrently)

Note(s): Not offered in 2016-17

**STAT 33580. Topics in Dynamical Systems: Exploring Chaotic Dynamics. 100 Units.**

This one-quarter dynamical systems topics course will focus on chaotic dynamical systems and their properties. The aim is for students to get a feel for properties associated with deterministic systems that exhibit chaotic behavior and to explore, through computational projects, how these are quantified. What is meant by "sensitive dependence on initial conditions" and how is this measured? How are correlations rapidly lost as nearby initial states evolve forward in time, and at what rate? How do we estimate an invariant measure on a chaotic attractor? What are typical "return times" in phase space, and how might we estimate their variance? What are generic properties of chaotic systems, and how can we understand these with simple paradigmatic constructions? What are generic mechanisms for creating chaotic dynamics by varying parameters of a dynamical system? This course investigates these questions through examples and takes an applied perspective.

Instructor(s): Mary Silber Terms Offered: Spring

Prerequisite(s): Consent of instructor



**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): Staff Terms Offered: Winter or Spring

Prerequisite(s): STAT 24500 or STAT 24510 is required; alternatively STAT 22400 and exposure to multivariate calculus. Some previous exposure to Fourier series is helpful but not required.

Equivalent Course(s): STAT 26100

**STAT 33610. Asymptotics for Time Series. 100 Units.**

This course will present a systematic asymptotic theory for time series analysis. In particular, the class will discuss asymptotics for sample mean, sample variances, banded covariance matrices estimates, inference of trends, periodograms, spectral density estimates, quantile estimation, nonparametric estimates, VaR and long-range dependent processes. Some asymptotic theory for non-stationary processes and functional linear models will also be presented.

Terms Offered: Autumn

Prerequisite(s): BUSF 30200 and STAT 31300 or consent of instructor

Note(s): Not offered in 2016-17

**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, co-integration and error-correction models, state-space models, dynamic factor models, and multivariate volatility models. The course also addresses impulse response function, structural specification, co-integration tests, least squares estimates, maximum likelihood estimates, structural changes, 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 package used in the course is R, but students may use their own software if preferred.

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>

Terms Offered: Spring

Prerequisite(s): BUSF 41910/STAT 33500

Equivalent Course(s): BUSF 41914

**STAT 33970. Statistics of High-Frequency Financial 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.

Instructor(s): P. Mykland Terms Offered: Winter

Prerequisite(s): STAT 39000/FINM 34500 (may be taken concurrently), also some statistics/econometrics background as in STAT 24400–24500, or FINM 33150 and FINM 33400, or equivalent, or consent of instructor.

Note(s): Not offered in 2016-17

Equivalent Course(s): FINM 33170

**STAT 34000. Gaussian Processes. 100 Units.**

Gaussian processes are commonly used in statistical models for spatial and spatial-temporal processes and for computer model output. They are also frequently used as building blocks for non-Gaussian process models. This course will begin with an overview of the theory for Gaussian processes, with a focus on stationary processes and their associated spectral properties and how these relate to problems of spatial interpolation. With this foundation, we will proceed to discuss a variety of approaches to developing useful classes of Gaussian process models, with a focus on spatial-temporal processes. Computational problems and possible solutions for fitting Gaussian process models to large, irregularly observed datasets will form the last part of the class. Applications to environmental monitoring data, computer model output and possibly other areas will be considered.

.This class is aimed at PhD students in Statistics, but may be accessible to others with a strong background in Statistics (say, STAT 24500 and 34300), some background in analysis and previous exposure to stochastic processes.

Terms Offered: Spring

Prerequisite(s): STAT 24500 and STAT 34300, or some background in analysis and previous exposure to stochastic processes

Note(s): Not offered in 2016-17

**STAT 34300. Applied Linear Statistical Methods. 100 Units.**

This course introduces the methods and applications of fitting and interpreting multiple regression models. The underlying distributional theory is discussed briefly. Topics include the examination of residuals, the transformation of data, strategies and criteria for the selection of a regression equation, and nonlinear models; categorical input variables (factors, constraints, and design matrices); factor models; factorial design; randomization; observational units versus experimental units; typology of experiments; randomized blocks design; and categorical responses (first case, logistic regression, likelihood analysis, and some basic asymptotic properties). The course emphasizes the use and interpretation of regression analysis with the R package. Techniques discussed are illustrated by examples involving both physical and social sciences data.

Terms Offered: Autumn

Prerequisite(s): Graduate student in Statistics or instructor consent

Note(s): Student who need it should take Linear Algebra (STAT 24300 or equivalent) concurrently.

**STAT 34700. Generalized Linear Models. 100 Units.**

This course covers exponential-family models; definition of generalized linear models; specific examples of GLMs; logistic and probit regression; cumulative logistic models; log-linear models and contingency tables; Quasi-likelihood and least squares; estimating functions; survival analysis; linear mixed models and generalized linear mixed models; and derivation of the methods are presented including likelihood analysis and some basic asymptotic properties. The course emphasizes the use and interpretation of generalized linear models with the R package. Techniques discussed are illustrated by examples involving physical, biological, and social science data.

Instructor(s): Staff Terms Offered: Winter

Prerequisite(s): STAT 34300 or consent of instructor

**STAT 34800. Graphical and Bayesian Models. 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.

Instructor(s): Staff Terms Offered: Spring

Prerequisite(s): STAT 34300 and STAT 34700 or consent of instructor

**STAT 34900. Data Analysis Project. 100 Units.**

The first half 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 individualized 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. It is intended that some projects in this class may develop into MS papers.

Instructor(s): M. Stein Terms Offered: Autumn

Prerequisite(s): STAT 34700 or 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): J. Dignam Terms Offered: Spring

Prerequisite(s): PBHS 32100 or STAT 22000; Introductory Statistics or Consent of Instructor

Equivalent Course(s): CCTS 32901, PBHS 32901

**STAT 35410. Genomic Evolution. 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 and J. Reinitz Terms Offered: Autumn

Equivalent Course(s): EVOL 35901, ECEV 35901

**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: Spring

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 30700 or PBHS 30900 or 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): 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: Spring. Not offered in 2017-18

Prerequisite(s): STAT 24400 and STAT 24500 or master level training in statistics.

Equivalent Course(s): PBHS 43010

**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): CHSS 32900, HIPS 25600, STAT 26700

**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: Autumn

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): 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.

Terms Offered: Autumn

Prerequisite(s): STAT 24400 is required; alternatively STAT 22400 and exposure to multivariate calculus and linear algebra.

Equivalent Course(s): STAT 27400

**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 Terms Offered: Spring

Prerequisite(s): CMSC 15400 or CMSC 12200 and STAT 22200 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.

Instructor(s): I. Kondor Terms Offered: Spring

Prerequisite(s): Consent of instructor

Equivalent Course(s): CMSC 35400

**STAT 37750. Compressed Sensing. 100 Units.**

The field of compressed sensing seeks to recover a high-dimensional signal from a relatively small number of observations. While impossible in general, in many settings this problem can be solved if  $x$  is sparse. Compressed sensing problems arise in countless applications, including image reconstruction, MRI, genetics, and many others. The course will also explore related questions such as different types of signal structure, and low-rank matrix completion (with applications to video denoising and to recommendation systems). This course will cover the theory and algorithms behind compressed sensing, as well as several applications. Students will apply these methods to real data sets as part of their homework. Prerequisites: familiar with linear algebra and probability; some programming experience is helpful but not required (the course will primarily use R or MATLAB).

Terms Offered: Spring

Prerequisite(s): STAT 30900. It is helpful but not required to have taken STAT 37601/37710/37790 or equivalent.

Note(s): Not offered in 2014-15

**STAT 37760. Modern Signal Processing. 100 Units.**

This course covers contemporary developments from time-frequency transforms and wavelets (1980s) to compressed sensing (2000s), a period during which signal processing significantly evolved and broadened to become the "mathematics of information". Topics: Review of classical sampling theory: Shannon-Nyquist, aliasing, filtering. Time-frequency transforms. Frame theory. Wavelet bases and filterbanks. Sparsity and nonlinear approximation. Algorithms: basis pursuit and matching pursuit. Compressed sensing. Matrix completion. Special topics: curvelets, phase retrieval, superresolution. Students who already have an interest in medical imaging (MRI, CT), or geophysical data processing (seismic, e-m), for instance, are welcome. The course assumes some affinity with undergraduate mathematics. The evaluation will consist of homework problems, and a project of the student's choice. The project can either consist in reproducing results from the literature, or can be research-oriented.

Terms Offered: Autumn

Prerequisite(s): Linear algebra and multivariate calculus

Note(s): Not offered in 2017-18

Equivalent Course(s): MATH 37760

**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: Autumn

Prerequisite(s): STAT 37710/CMSC 35400 or consent of instructor

Note(s): Not offered in 2017-18

**STAT 37810. Statistical Computing A. 050 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 37820. Statistical Computing B. 050 Units.**

Statistical Computing B 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 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  $L_p$  spaces, Radon-Nikodym theorem, conditional expectation, and martingale theory.

Terms Offered: Spring

Prerequisite(s): STAT 38100

**STAT 38500. 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: Autumn

Prerequisite(s): STAT 38300 or MATH 31200, or permission of the instructor.

Equivalent Course(s): MATH 38509

**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.

Instructor(s): G. Lawler Terms Offered: Autumn

Prerequisite(s): The usual prerequisites are either the first-year graduate mathematical analysis sequence (mainly the material in MATH 31200) or STAT 38100-38300, the first two quarters of the statistics measure-theoretic probability sequence.

Equivalent Course(s): MATH 38511

**STAT 38600. Topics in Stochastic Processes. 100 Units.**

This will be a course in "high-dimensional" probability aimed at introducing some of the mathematics of empirical processes, concentration, Gaussian random fields, large random matrices, and compressed sensing.

Terms Offered: TBD

Prerequisite(s): Basic probability and analysis, discrete-time martingales (STAT 30400 and 31300)

Note(s): Not offered in 2016-17

**STAT 38620. Social Networks, Probability, Learning, and Game Theory. 100 Units.**

This is a research oriented topic course aimed at graduate students. We will first cover some basics of social networks including structure and analysis of such networks and models that abstract their basic properties. Then we will focus on some recent research on a few selected topics/models, and aim to discuss one representative example in each of the following topics: (1) Probabilistic models and statistical learning based on empirical observation; (2) Stochastic processes (such as spread of information) and game-theoretical behavior on social networks as well as corresponding optimization problems; (3) Connections with social choices relating to collective decision making; (4) Some algorithmic aspects of networks. The students should have solid knowledge in at least two of the following areas: (1) Probability theory (either 31200-31300 or 38100-38300). (2) Statistics (either 24400-24500-24610 or 30400-30100-30210). (3) Basic knowledge in game theory and algorithms. In addition, students should be comfortable with undergraduate linear algebra as well as elementary combinatorics.

Terms Offered: Winter

Prerequisite(s): Consent of instructor. Students need to be familiar with two out of the following three: probability (no need for measure theory)/statistics/game theory (at intro level).

Note(s): Not offered in 2014-15

**STAT 38650. Random Matrices and Related Topics. 100 Units.**

This course will be an introduction to the spectral theory of large random matrices and related topics in probability. The first part of the course will be devoted to bulk spectral properties of Wigner and sample covariance matrices (that is, the empirical distribution of their eigenvalues), leading to the Wigner semi-circle law and the Marchenko-Pastur theorem. The second part will focus on the Gaussian orthogonal and unitary ensembles and on the distribution theory of the top eigenvalue (Tracy-Widom theory). This will lead to the study of orthogonal polynomials, Fredholm determinants, determinantal point processes, and Toeplitz matrices. Relationships to various combinatorial problems in probability, including asymmetric exclusion processes, last-passage percolation, and various stochastic models of growth and deposition, will be studied. Several other related topics may be discussed, depending on the interests and backgrounds of the audience and the instructor.

Note(s): Not offered in 2016-17

**STAT 38660. Random Planar Geometry. 100 Units.**

This is a research topic course on certain aspects of random planar geometry. The two central models to be discussed are Liouville quantum gravity which arises from exponentiating a two-dimensional Gaussian free field, as well as uniform infinite planar triangulation/quadrangulation. We will mainly focus on the discrete perspectives of these models, but will also at times discuss the connections to the continuous counterparts. We will concentrate on the metric properties of these random surfaces (including geodesic distances and the electric resistances), as well as their connections to the random motion on these random surfaces.

Terms Offered: Autumn

Prerequisite(s): Recommended 38100/38300 sequence, or experience with measure-theoretical probability.

Note(s): Not offered in 2016-17

**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 39800. Field Research. Variable 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): Consent of instructor and faculty advisor

**STAT 39900. Master's Seminar. Variable 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.

**STAT 40100. Reading/Research: Statistics. Variable 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): 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 mis-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 mis-specified models. The courses may be taken separately.

Terms Offered: Autumn

Prerequisite(s): STAT 30100 and STAT 30400 and STAT 31015, or consent of instructor

**STAT 41600. High-Dimensional Statistics II. 100 Units.**

No description available.

Terms Offered: Spring

Prerequisite(s): STAT 30100 or STAT 30400 or STAT 31015, or consent of instructor

**STAT 42510. Theoretical Neuroscience: Single Neuron Dynamics and Computation. 100 Units.**

This course is the first part of a three-quarter sequence in theoretical/computational neuroscience. It will focus on mathematical models of single neurons. Topics will include: basic biophysical properties of neurons; Hodgkin-Huxley model for action potential generation; 2D models, phase-plane analysis and bifurcations leading to action potential generation; integrate-and-fire-type models; noise; characterization of neuronal activity with stochastic inputs; spatially extended models; models of synaptic currents and synaptic plasticity; unsupervised learning; supervised learning; reinforcement learning.

Terms Offered: Autumn

Prerequisite(s): Prior exposure to differential equations, linear algebra, probability theory

Equivalent Course(s): CPNS 35510

**STAT 42520. Theoretical Neuroscience: Network Dynamics and Computation. 100 Units.**

This course is the second part of a three-quarter sequence in theoretical/computational neuroscience. It will focus on mathematical models of networks of neurons. Topics will include: firing rate models for populations of neurons; spatially extended firing rate models; models of visual cortex; models of brain networks at different levels; characterization of properties of specific brain networks; models of networks of binary neurons, mean rates, correlations, reductions to rate models; learning in networks of binary neurons, associative memory models; models of networks of spiking neurons: asynchronous vs synchronous states; oscillations in networks of spiking neurons; learning in networks of spiking neurons; models of working memory; models of decision-making.

Terms Offered: Winter

Prerequisite(s): Prior exposure to differential equations, linear algebra, probability theory, STAT 42510 or instructor consent.

Equivalent Course(s): CPNS 35520

**STAT 42600. Theoretical Neuroscience: Statistics and Information Theory. 100 Units.**

This course is the third part of a three-quarter sequence in theoretical/computational neuroscience. It begins with the spike sorting problem, used as 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 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 and newer population coding models. The decoding section will cover basic methods for inferring the stimulus 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 45800. Workshop on Collaborative Research in Statistics, Computing, and Science. 100 Units.**

This course aims to bring together researchers with expertise in statistics, computation, and basic sciences, to work together to produce a solution to a particular problem. The problem we will focus on is the following: how can we improve the way that statistical comparisons are performed? No knowledge of this problem is assumed: it will be introduced in full at the start of the class, together with an outline for an initial proposed approach to addressing the problem. In brief the motivation is as follows:

Many new statistical methods are published without any software implementation, and without any comparisons with existing methods. Even when comparisons are made, usually the comparisons are performed by a single research group who has developed one of the methods, raising the concern that the comparison may unfairly favor this method. Indeed, this problem is almost inevitable, even if the authors are extremely fastidious: any research group will have different levels of expertise with different methods, and tend to be more effective in applying their own method. Indeed, getting a method to work well for a particular problem may in itself be a research project. On top of this, performing these kinds of comparisons is incredibly time-consuming: at a minimum one has to familiarize oneself with a range of software products, their input/output requirements, and their various run-time options; create an infrastructure for running them; and write scripts to compare the

Terms Offered: Winter

Prerequisite(s): Consent of instructor

Note(s): Note offered in 2016-17

**STAT 48100. Proseminar in Probability. 100 Units.**

This course will explore topics of current research interest in probability theory and stochastic processes. Students will be expected to give presentations based on research articles chosen after consultation with the instructors.

Instructor(s): Steven Lalley, Staff Terms Offered: Autumn, Spring, Winter

Prerequisite(s): Consent of instructor





#### Font Notice

This document should contain certain fonts with restrictive licenses. For this draft, substitutions were made using less legally restrictive fonts. Specifically:

Times was used instead of Trajan.

Times was used instead of Palatino.

The editor may contact Leepfrog for a draft with the correct fonts in place.