The department’s goals are to acquaint students with the role played in science and technology by probabilistic and statistical ideas and methods, to provide instruction in the theory and application of techniques that have been found to be commonly useful, and to train research workers in probability and statistics. There are courses for general students as well as those who plan careers in statistics in business, government, industry, and teaching.

The department has long recognized the relation of statistical theory to applications. It has fostered this by encouraging a liaison with other departments in the form of joint and courtesy faculty appointments, as well as membership in various interdisciplinary programs: Biomedical Data Science, Bio-X, Center for Computational, Evolutionary and Human Genomics, Computer Science, Economics, Education, Electrical Engineering, Environmental Earth System Science, Genetics, Mathematics, Mathematical and Computational Finance, and Medicine. The research activities of the department reflect an interest in applied and theoretical statistics and probability. There are workshops in biology/medicine and in environmental factors in health.

In addition to courses for Statistics students, the department offers a number of service courses designed for students in other departments. These tend to emphasize the application of statistical techniques rather than their theoretical development.

The department has always drawn visitors from other countries and universities, and as a result there are a wide range of seminars offered by both the visitors and the department’s own faculty.

Undergraduate Programs in Statistics
The department offers a minor in Statistics. More information can be found under the Minor section.

Undergraduates Interested in Statistics
Students wishing to build a concentration in probability and statistics are encouraged to consider declaring a major in Mathematical and Computational Science (https://mcs.stanford.edu). This interdepartmental program is administered in the Department of Statistics and provides core training in computing, mathematics, operations research, and statistics, with opportunities for further elective work and specialization. See the "Mathematical and Computational Science" section of this bulletin.

Graduate Programs in Statistics
University requirements for the M.S. and Ph.D. degrees are discussed in the "Graduate Degrees (http://exploredegrees.stanford.edu/archive/2016-17/graduatedegrees)" section of this bulletin.

Learning Outcomes (Graduate)
The purpose of the master’s program is to further develop knowledge and skills in Statistics and to prepare students for a professional career or doctoral studies. This is achieved through completion of courses, in the primary field as well as related areas, and experience with independent work and specialization.

The Ph.D. is conferred upon candidates who have demonstrated substantial scholarship and the ability to conduct independent research and analysis in Statistics. Through completion of advanced course work and rigorous skills training, the doctoral program prepares students to make original contributions to the knowledge of Statistics and to interpret and present the results of such research.

Minor in Statistics
The undergraduate minor in Statistics is designed to complement major degree programs primarily in the social and natural sciences. Students with an undergraduate Statistics minor should find broadened possibilities for employment. The Statistics minor provides valued preparation for professional degree studies in postgraduate academic programs.

The minor consists of a minimum of six courses with a total of at least 20 units. There are two required courses (8 units) and four qualifying or elective courses (12 or more units). All courses for the minor must be taken for a letter grade. An overall 2.75 grade point average (GPA) is required for courses fulfilling the minor.

Required Courses
<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATS 116</td>
<td>Theory of Probability</td>
<td>3-5</td>
</tr>
<tr>
<td>STATS 200</td>
<td>Introduction to Statistical Inference</td>
<td>3</td>
</tr>
</tbody>
</table>

Qualifying Courses
At most, one of these two courses may be counted toward the six course requirement for the minor:

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATH 52</td>
<td>Integral Calculus of Several Variables</td>
<td>5</td>
</tr>
<tr>
<td>STATS 191</td>
<td>Introduction to Applied Statistics</td>
<td>3-4</td>
</tr>
</tbody>
</table>

Elective Courses
At least one of the elective courses should be a STATS 200-level course. The remaining two elective courses may also be 200-level courses. Alternatively, one or two elective courses may be approved courses in other departments. Special topics courses and seminars for undergraduates are offered from time to time by the department, and these may be counted toward the course requirement. Students may not count any Statistics courses below the 100 level toward the minor.

Examples of elective course sequences are:

<table>
<thead>
<tr>
<th>Course Sequence</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA 202 Data Mining and Analysis</td>
<td>3</td>
</tr>
<tr>
<td>DATA 203 Introduction to Regression Models and Analysis of Variance</td>
<td>3</td>
</tr>
</tbody>
</table>

Statistical Methodology

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATS 205</td>
<td>Introduction to Nonparametric Statistics</td>
<td>3</td>
</tr>
<tr>
<td>STATS 206</td>
<td>Applied Multivariate Analysis</td>
<td>3</td>
</tr>
<tr>
<td>STATS 207</td>
<td>Introduction to Time Series Analysis</td>
<td>3</td>
</tr>
</tbody>
</table>

Economic Optimization

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATS 206</td>
<td>Applied Multivariate Analysis</td>
<td>3</td>
</tr>
<tr>
<td>ECON 160</td>
<td>Game Theory and Economic Applications</td>
<td>5</td>
</tr>
</tbody>
</table>

Psychology Modeling and Experiments

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATS 206</td>
<td>Applied Multivariate Analysis</td>
<td>3</td>
</tr>
</tbody>
</table>

Signal Processing

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATS 207</td>
<td>Introduction to Time Series Analysis</td>
<td>3</td>
</tr>
<tr>
<td>EE 264</td>
<td>Digital Signal Processing</td>
<td>3</td>
</tr>
<tr>
<td>EE 279</td>
<td>Introduction to Digital Communication</td>
<td>3</td>
</tr>
</tbody>
</table>

Genetic and Ecologic Modeling

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATS 217</td>
<td>Introduction to Stochastic Processes I</td>
<td>3</td>
</tr>
<tr>
<td>BIO 283</td>
<td>Theoretical Population Genetics</td>
<td>3</td>
</tr>
</tbody>
</table>
the graduate career. No courses taken prior to the first quarter of the sophomore year may be used to meet master’s degree requirements.

Course transfers are not possible after the bachelor’s degree has been conferred.

The University requires that the graduate adviser be assigned in the student’s first graduate quarter even though the undergraduate career may still be open. The University also requires that the Master’s Degree Program Proposal be completed by the student and approved by the department by the end of the student’s first graduate quarter.

Master of Science in Statistics
Curriculum and Degree Requirements
The department requires that a master’s student take 45 units of work from offerings in the Department of Statistics (https://exploredegrees.stanford.edu/search?view=catalog&filter-coursestatus=Active=on&page=0&catalog=&academicYear=&q=STATS&collapse=) or from authorized courses in other departments. With the advice of the master’s program advisers, each student selects his or her own set of electives.

All requirements for the Statistics master’s degree, including the coterminal master’s degree, must be completed within three years of the first quarter of graduate standing. Ordinarily, four or five quarters are needed to complete all requirements. Honors Cooperative students must finish within five years.

Units for a given course may not be counted to meet the requirements of more than one degree, with the exception that up to 45 units of a Stanford M.A. or M.S. degree may be applied to the residency requirement for the Ph.D., D.M.A. or Engineer degrees. See the “Residency Policy for Graduate Students (http://exploredegrees.stanford.edu/archive/2016-17/graduatedegrees/#residencytext)” section of this Bulletin for University rules.

As defined in the general graduate student requirements, students must maintain a grade point average (GPA) of 3.0 or better and classes must be taken at the 200 level or higher. No thesis is required.


1. Statistics Core Courses (must complete all four courses):

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATS 116</td>
<td>Theory of Probability</td>
<td>3-5</td>
</tr>
<tr>
<td>STATS 191</td>
<td>Introduction to Applied Statistics</td>
<td>3-4</td>
</tr>
<tr>
<td>STATS 200</td>
<td>Introduction to Statistical Inference</td>
<td>3</td>
</tr>
<tr>
<td>STATS 217</td>
<td>Introduction to Stochastic Processes I</td>
<td>2-3</td>
</tr>
</tbody>
</table>

All must be taken for a letter grade, with the exception of courses offered satisfactory/no credit only. Students with prior background may replace each course with a more advanced course from the same area. Courses previously taken may be waived by the adviser, in which case they must be replaced by other graduate courses offered by the department.

2. Linear Algebra Mathematics Requirement:

Select one of the following:

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATH 104</td>
<td>Applied Matrix Theory</td>
<td>3</td>
</tr>
<tr>
<td>MATH 113</td>
<td>Linear Algebra and Matrix Theory</td>
<td>3</td>
</tr>
<tr>
<td>MATH 115</td>
<td>Functions of a Real Variable</td>
<td>3</td>
</tr>
<tr>
<td>MATH 171</td>
<td>Fundamental Concepts of Analysis</td>
<td>3</td>
</tr>
</tbody>
</table>
All must be taken for a letter grade, with the exception of courses offered satisfactory/no credit only. Substitution of other courses in Mathematics may be made with consent of the adviser.

3. Programming Requirement:

Select one of the following:

- CS 106A Programming Methodology 3
- CS 106B Programming Abstractions 3
- CS 106X Programming Abstractions (Accelerated) 3
- CME 108 Introduction to Scientific Computing 3

All must be taken for a letter grade, with the exception of courses offered satisfactory/no credit only. Substitution of other courses in Computer Science may be made with consent of the adviser.

4. Additional Statistics Courses:

At least four additional Statistics courses must be taken from graduate offerings in the department (STATS 202 through 390). All must be taken for a letter grade, with the exception of courses offered satisfactory/no credit only.

Students cannot count more than a total 6 units of the following toward the master's degree requirements:

- STATS 260A Workshop in Biostatistics 1-2
- STATS 260B Workshop in Biostatistics 1-2
- STATS 260C Workshop in Biostatistics 1-2
- STATS 298 Industrial Research for Statisticians 1
- STATS 299 Independent Study 1-10
- STATS 390 Consulting Workshop 1

5. Elective Courses:

Additional elective units to complete the requirements may be chosen from the list available from the department web site (https://statistics.stanford.edu/academics/ms-statistics-elective-courses) . Other graduate courses (200 or above) may be authorized by the adviser if they provide skills relevant to degree requirements or deal primarily with an application of statistics or probability and do not overlap courses in the student’s program. There is sufficient flexibility to accommodate students with interests in applications to business, computing, economics, engineering, health, operations research, and biological and social sciences.

Courses below 200 level are not acceptable, with the following exceptions:

- STATS 116 Theory of Probability 3-5
- STATS 191 Introduction to Applied Statistics 3-4
- MATH 104 Applied Matrix Theory 3
- MATH 113 Linear Algebra and Matrix Theory 3
- MATH 115 Functions of a Real Variable 3
- MATH 171 Fundamental Concepts of Analysis 3
- CS 106A Programming Methodology 3-5
- CS 106B Programming Abstractions 3-5
- CS 106X Programming Abstractions (Accelerated) 3-5
- CS 140 Operating Systems and Systems Programming 3-4
- CS 142 Web Applications 3
- CS 143 Compilers 3-4
- CS 144 Introduction to Computer Networking 3-4
- CS 145 Introduction to Databases 3-4
- CS 147 Introduction to Human-Computer Interaction Design 3-5

At most, one of these courses may be counted:

- MATH 104 Applied Matrix Theory 3
- MATH 113 Linear Algebra and Matrix Theory 3
- STATS 116 Theory of Probability 3-5

6. Master's Degree Program Proposal

The Statistics Master's Degree Program Proposal form (https://statistics.stanford.edu/masters-program-proposal-form) , signed and approved by the student’s program adviser, must be submitted by the student to the major department’s student services administrator prior to the end of the first quarter of enrollment in the program. A revised program proposal must be submitted if degree plans change.

There is no thesis requirement.

Students with a strong mathematical background who are interested in pursuing a Ph.D. in Statistics should consider applying to the Ph.D. program.

Master of Science in Statistics: Data Science

The Data Science track develops strong mathematical, statistical, and computational and programming skills through the general master’s core and programming requirements. In addition, it provides a fundamental data science education through general and focused electives requirement from courses in data sciences and related areas. Course choices are limited to predefined courses from the data sciences and related courses group. Programming requirement (requirement 4) is extended to 6 units and includes course work in advanced scientific programming and high performance computing. The final requirement is a practical component (requirement 5) for 6 units to be completed through capstone project, data science clinic, or other courses that have strong hands-on or practical component, such as statistical consulting.

Admission

Prospective applicants should consult the Graduate Admissions (https://studentaffairs.stanford.edu/gradadmissions) and the Statistics Department admissions webpages (https://statistics.stanford.edu/academics/admissions) for complete information on admission requirements and deadlines.

Applicants apply to the Master of Science degree in Statistics and declare preference for the Data Science subplan within the application ("Department Specialization" option). Selection of the students is made by the Statistics admission committee, which has representation from the Data Science steering committee.

Prerequisites

Fundamental courses in mathematics and computing may be needed as prerequisites for other courses in the program. Check the prerequisites of each required course. Recommended preparatory courses include advanced undergraduate level courses in linear algebra, probability,
and introductory courses in PDEs, stochastics, numerical methods and proficiency in programming.

Curriculum and Degree Requirements

As defined in the general graduate student requirements, students must maintain a grade point average (GPA) of 3.0 or better and classes must be taken at the 200 level or higher. Students must complete 45 units of required coursework in Data Science.

A Master's Degree Program Proposal (https://statistics.stanford.edu/stats-ds-program-proposa-form-pdf), signed and approved by the student's program adviser, is to be submitted by the student to the major department's student services administrator prior to the end of the first quarter of enrollment in the program. A revised program proposal must be submitted if degree plans change.

The Data Science subplan is printed on the transcript and diploma. No thesis is required.

Requirement 1: Foundational (12 units)

Students must demonstrate foundational knowledge in the field by completing the following core courses. Courses in this area must be taken for letter grades.

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>CME 302</td>
<td>Numerical Linear Algebra</td>
<td>3</td>
</tr>
<tr>
<td>CME 305</td>
<td>Discrete Mathematics and Algorithms</td>
<td>3</td>
</tr>
<tr>
<td>CME 307</td>
<td>Optimization</td>
<td>3</td>
</tr>
<tr>
<td>CME 308</td>
<td>Stochastic Methods in Engineering</td>
<td>3</td>
</tr>
<tr>
<td>CME 309</td>
<td>Randomized Algorithms and Probabilistic Analysis</td>
<td>3</td>
</tr>
</tbody>
</table>

Requirement 2: Data Science Electives (12 units)

Data Science electives should demonstrate breadth of knowledge in the technical area. The elective course list is defined. Courses outside this list are subject to approval. Courses in this area must be taken for letter grades.

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATS 200</td>
<td>Introduction to Statistical Inference</td>
<td>3</td>
</tr>
<tr>
<td>STATS 203</td>
<td>Introduction to Regression Models and Analysis of Variance</td>
<td>3</td>
</tr>
<tr>
<td>or STATS 305A</td>
<td>Introduction to Statistical Modeling</td>
<td>3</td>
</tr>
<tr>
<td>STATS 315A</td>
<td>Modern Applied Statistics: Learning</td>
<td>2-3</td>
</tr>
<tr>
<td>STATS 315B</td>
<td>Modern Applied Statistics: Data Mining</td>
<td>2-3</td>
</tr>
</tbody>
</table>

Requirement 3: Specialized Electives (9 units)

Choose three courses in specialized areas from the following list. Courses outside this list are subject to approval.

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI O 214</td>
<td>Representations and Algorithms for Computational Biology</td>
<td>3-4</td>
</tr>
<tr>
<td>BIomedIN 215</td>
<td>Data Driven Medicine</td>
<td>3</td>
</tr>
<tr>
<td>BIOS 221/</td>
<td>Modern Statistics for Modern Biology</td>
<td>3</td>
</tr>
<tr>
<td>STATS 366</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS 224W</td>
<td>Social and Information Network Analysis</td>
<td>3-4</td>
</tr>
<tr>
<td>CS 229</td>
<td>Machine Learning</td>
<td>3-4</td>
</tr>
<tr>
<td>CS 246</td>
<td>Mining Massive Data Sets</td>
<td>3-4</td>
</tr>
<tr>
<td>CS 347</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>CS 448</td>
<td>Topics in Computer Graphics</td>
<td>3-4</td>
</tr>
<tr>
<td>ENERGY 240</td>
<td>Geostatistics</td>
<td>2-3</td>
</tr>
<tr>
<td>OIT 367</td>
<td>Business Intelligence from Big Data</td>
<td>3</td>
</tr>
<tr>
<td>PSYCH 204A</td>
<td>Human Neuroimaging Methods</td>
<td></td>
</tr>
<tr>
<td>STATS 290</td>
<td>Paradigms for Computing with Data</td>
<td>3</td>
</tr>
</tbody>
</table>

Requirement 4: Advanced Scientific Programming and High Performance Computing Core (6 units)

To ensure that students have a strong foundation in programming, 3 units of advanced scientific programming for letter grade at the level of CME212 and three units of parallel computing for letter grades are required.

Note: Programming proficiency at the level of CME 211 is a hard prerequisite for CME 212 (students may only place out of 211 with prior written approval). CME 211 can be applied towards elective requirement.

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>CME 212</td>
<td>Advanced Software Development for Scientists and Engineers</td>
<td>3</td>
</tr>
<tr>
<td>Parallel Computing/HCP courses: (3 units)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CME 213</td>
<td>Introduction to parallel computing using MPI, openMP, and CUDA</td>
<td>3</td>
</tr>
<tr>
<td>CME 323</td>
<td>Distributed Algorithms and Optimization</td>
<td>3</td>
</tr>
<tr>
<td>CME 342</td>
<td>Parallel Methods in Numerical Analysis</td>
<td>3</td>
</tr>
<tr>
<td>CS 149</td>
<td>Parallel Computing</td>
<td>3-4</td>
</tr>
<tr>
<td>CS 315A</td>
<td>Parallel Computer Architecture and Programming</td>
<td>3</td>
</tr>
<tr>
<td>CS 316</td>
<td>Advanced Multi-Core Systems</td>
<td>3</td>
</tr>
<tr>
<td>CS 344C, offered in previous years, may also be counted</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Requirement 5: Practical Component (6 units)

Students are required to take 6 units of practical component that may include any combination of:

- A capstone project, supervised by a faculty member and approved by the student's adviser. The capstone project should be computational in nature. Students should submit a one-page proposal, supported by the faculty member and sent to the student's Data Science adviser for approval (at least one quarter prior to start of project).
- Master's research: STATS 299 Independent Study.
- Project labs offered by Stanford Data Lab: ENGR 250 Data Challenge Lab , and ENGR 350 Data Impact Lab .
- Other courses that have a strong hands-on and practical component, such as STATS 390 Consulting Workshop up to 1 unit.

Doctor of Philosophy in Statistics

The department looks for students who wish to prepare for research careers in statistics or probability, either applied or theoretical. Advanced undergraduate or master's level work in mathematics and statistics provides a good background for the doctoral program. Quantitatively oriented students with degrees in other scientific fields are also encouraged to apply for admission. The program normally takes five years to complete.

Program Summary

<table>
<thead>
<tr>
<th>Course</th>
<th>Title</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-year core program</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STATS 300</td>
<td>Advanced Topics in Statistics: Stochastic Block Models and Latent Variable Models (offered Summer Quarter)</td>
<td>2-3</td>
</tr>
<tr>
<td>STATS 300A</td>
<td>Theory of Statistics I</td>
<td>2-3</td>
</tr>
<tr>
<td>STATS 300B</td>
<td>Theory of Statistics II</td>
<td>2-4</td>
</tr>
</tbody>
</table>
Students are expected to show acceptable performance in two theory and methods, applied statistics, and probability theory, and all are typically taken during the summer between the student’s first and second years. Students are required to take 15 units of coursework outside of the department and are advised to choose an area of concentration in a specific scientific field of statistical applications approved by their Ph.D. program adviser.


Dissertation Reading Committee, Dissertation Proposal Meeting and University Oral Examinations

The dissertation reading committee consists of the student’s adviser plus two faculty readers, all of whom are responsible for reading and approving the full dissertation.

The dissertation proposal meeting is intended to demonstrate students’ depth in some areas of statistics, and to examine the general plan for their research. It also confirms that students have chosen a Ph.D. faculty adviser and have started to work with that adviser on a research topic. In the meeting, they will give a short presentation and discuss their ideas for completing a Ph.D. thesis, with a committee typically consisting of the members of the dissertation reading committee. The meeting must be successfully completed before the end of the third year. "Successful completion" means that the general research plan is sound and has a reasonable chance of success. If the student does not pass, the meeting must be repeated. Repeated failure by the end of Year 3 can lead to a loss of financial support.

Nearly all students can expect to pass this examination, although it is common for specific recommendations to be made regarding completion of the written dissertation.

For further information on University oral examinations and committees, see the Graduate Academic Policies and Procedures (GAP) Handbook, section 4.7 (http://gap.stanford.edu/4-7.html) or the "University Oral Examination (http://explorerdegrees.stanford.edu/archive/2016-17/graduatedegrees/#doctoraltext) " section of this bulletin.

Doctoral and Research Advisers

From the student’s arrival until the selection of a research adviser, the student’s academic progress is monitored by the department’s Director of Graduate Studies. Each student should meet at least once a quarter with the Doctoral Adviser to discuss their academic plans and their progress towards choosing a dissertation adviser.

Financial Support

Students accepted to the Ph.D. program are offered financial support. All tuition expenses are paid and there is a fixed monthly stipend determined to be sufficient to pay living expenses. Financial support can be continued for five years, department resources permitting, for students who do not have enough mathematics background can take STATS 310A, B, C after their first year but need to have their first-year program approved by the Director of Graduate Studies.

Qualifying Examinations

These are intended to test the student’s level of knowledge when the first-year program, common to all students, has been completed. There are separate examinations in the three core subjects of statistical theory and methods, applied statistics, and probability theory, and all are typically taken during the summer between the student’s first and second years. Students are expected to show acceptable performance in two examinations. Letter grades are not given. After passing the qualifying exams students file for Ph.D. candidacy, a University milestone.
in good standing. The resources for student financial support derive from funds made available for student teaching and research assistantships. Students receive both a teaching and research assignment each quarter which, together, do not exceed 20 hours. Students are encouraged to apply for outside scholarships, fellowships, and other forms of financial support.

**Ph.D. Minor in Statistics**

Students must complete 30 total units for the Ph.D. minor. 20 units must be from Statistics courses numbered 300 and above and taken for a letter grade (minimum grade of B for each course). The remaining 10 units can be from Statistics courses numbered 200 and above, and may be taken for credit. Students may not include more than three units of Stats 390, Consulting Workshop, towards the 30 units. The selection of courses must be approved by the Director of Graduate Studies. The Application for the Ph.D. Minor form must be approved by both the student's Ph.D. department and the Statistics department.

For further information about the Statistics Ph.D. degree program requirements, see the department web site (https://statistics.stanford.edu/academics/doctoral-program).

_Emeriti: (Professors) Jerome H. Friedman, Charles Stein, Paul Switzer; (Courtesy Professor) Richard A. Olshen_

Chair: Emmanuel Candès


_Associate Professor:_ Chiara Sabatti

_Assistant Professors:_ John Duchi, Julia Palacios, Balakanapathy Rajaratnam

_Courtesy Professors:_ John Ioannidis, Philip W. Lavori

_Courtesy Associate Professors:_ David Rogosa, Hua Tang, Lu Tian

_Courtesy Assistant Professors:_ Mike Baiocchi, Percy Shuo Liang

_Consulting Professor:_ John Chambers

_Stein Fellows:_ James Johndrow, Rajarshi Mukherjee, Rachel Wang, Lucy Xia

**Courses**

**STATS 42Q. Undergraduate Admissions to Selective Universities - a Statistical Perspective. 2 Units.**

The goal is the building of a statistical model, based on applicant data, for predicting admission to selective universities. The model will consider factors such as gender, ethnicity, legacy status, public-private schooling, test scores, effects of early action, and athletics. Common misconceptions and statistical pitfalls are investigated. The applicant data are not those associated with any specific university.

**STATS 48N. Riding the Data Wave. 3 Units.**

Imagine collecting a bit of your saliva and sending it in to one of the personalized genomics company: for very little money you will get back information about hundreds of thousands of variable sites in your genome. Records of exposure to a variety of chemicals in the areas you have lived are only a few clicks away on the web; as are thousands of studies and informal reports on the effects of different diets, to which you can compare your own. What does this all mean for you? Never before in history humans have recorded so much information about themselves and the world that surrounds them. Nor has this data been so readily available to the lay person. Expression as “data deluge” are used to describe such wealth as well as the loss of proper bearings that it often generates. How to summarize all this information in a useful way? How to boil down millions of numbers to just a meaningful few? How to convey the gist of the story in a picture without misleading oversimplifications? To answer these questions we need to consider the use of the data, appreciate the diversity that they represent, and understand how people instinctively interpret numbers and pictures. During each week, we will consider a different data set to be summarized with a different goal. We will review analysis of similar problems carried out in the past and explore if and how the same tools can be useful today. We will pay attention to contemporary media (newspapers, blogs, etc.) to identify settings similar to the ones we are examining and critique the displays and summaries there documented. Taking an experimental approach, we will evaluate the effectiveness of different data summaries in conveying the desired information by testing them on subsets of the enrolled students.

**STATS 50. Mathematics of Sports. 3 Units.**


Same as: MCS 100

**STATS 60. Introduction to Statistical Methods: Precalculus. 5 Units.**

Techniques for organizing data, computing, and interpreting measures of central tendency, variability, and association. Estimation, confidence intervals, tests of hypotheses, t-tests, correlation, and regression. Possible topics: analysis of variance and chi-square tests, computer statistical packages.

Same as: PSYCH 10, STATS 160

**STATS 101. Data Science 101. 5 Units.**

This course will provide a hands-on introduction to statistics and data science. Students will engage with the fundamental ideas in inferential and computational thinking. Each week, we will explore a core topic comprising three lectures and two labs (a module), in which students will manipulate real-world data and learn about statistical and computational tools. Students will engage in statistical computing and visualization with current data analytic software (Jupyter, R). The objectives of this course are to have students (1) be able to connect data to underlying phenomena and to think critically about conclusions drawn from data analysis, and (2) be knowledgeable about programming abstractions so that they can later design their own computational inferential procedures. No programming or statistical background is assumed. Freshmen and sophomores interested in data science, computing and statistics are encouraged to attend. Open to graduates as well. http://web.stanford.edu/class/stats101/.

**STATS 110. Statistical Methods in Engineering and the Physical Sciences. 4-5 Units.**

Introduction to statistics for engineers and physical scientists. Topics: descriptive statistics, probability, interval estimation, tests of hypotheses, nonparametric methods, linear regression, analysis of variance, elementary experimental design. Prerequisite: one year of calculus.
STATS 116. Theory of Probability. 3-5 Units.
Probability spaces as models for phenomena with statistical regularity. Discrete spaces (binomial, hypergeometric, Poisson). Continuous spaces (normal, exponential) and densities. Random variables, expectation, independence, conditional probability. Introduction to the laws of large numbers and central limit theorem. Prerequisites: MATH 52 and familiarity with infinite series, or equivalent.

STATS 141. Biostatistics. 3-5 Units.
Introductory statistical methods for biological data: describing data (numerical and graphical summaries); introduction to probability; and statistical inference (hypothesis tests and confidence intervals). Intermediate statistical methods: comparing groups (analysis of variance); analyzing associations (linear and logistic regression); and methods for categorical data (contingency tables and odds ratio). Course content integrated with statistical computing in R.
Same as: BIO 141

STATS 155. Statistical Methods in Computational Genetics. 3 Units.
The computational methods necessary for the construction and evaluation of sequence alignments and phylogenies built from molecular data and genetic data such as micro-arrays and data base searches. How to formulate biological problems in an algorithmic decomposed form, and building blocks common to many problems such as Markovian models, multivariate analyses. Some software covered in labs (Python, Biopython, XGobi, MrBayes, HMMER, Probe). Prerequisites: knowledge of probability equivalent to STATS 116, STATS 202 and one class in computing at the CS 106 level. Writing intensive course for undergraduates only. Instructor consent required. (WIM).

STATS 160. Introduction to Statistical Methods: Precalculus. 5 Units.
Techniques for organizing data, computing, and interpreting measures of central tendency, variability, and association. Estimation, confidence intervals, tests of hypotheses, t-tests, correlation, and regression. Possible topics: analysis of variance and chi-square tests, computer statistical packages.
Same as: PSYCH 10, STATS 60

STATS 167. Probability: Ten Great Ideas About Chance. 4 Units.
Foundational approaches to thinking about chance in matters such as gambling, the law, and everyday affairs. Topics include: chance and decisions; the mathematics of chance; frequencies, symmetry, and chance; Bayes great idea; chance and psychology; misuses of chance; and harnessing chance. Emphasis is on the philosophical underpinnings and problems. Prerequisite: exposure to probability or a first course in statistics at the level of STATS 60 or 116.
Same as: PHIL 166, PHIL 266, STATS 267

STATS 191. Introduction to Applied Statistics. 3-4 Units.
Statistical tools for modern data analysis. Topics include regression and prediction, elements of the analysis of variance, bootstrap, and cross-validation. Emphasis is on conceptual rather than theoretical understanding. Applications to social/biological sciences. Student assignments/projects require use of the software package R. Recommended: 60, 110, or 141.

STATS 195. Introduction to R. 1 Unit.
This short course runs for the first four weeks of the quarter and is offered in fall and spring. It is recommended for students who want to use R in statistics, science, or engineering courses and for students who want to learn the basics of R programming. The goal of the short course is to familiarize students with R's tools for scientific computing. Lectures will be interactive with a focus on learning by example, and assignments will be application-driven. No prior programming experience is needed. Topics covered include basic data structures, File I/O, graphs, control structures, etc, and some useful packages in R.
Same as: CME 195

STATS 196A. Multilevel Modeling Using R. 1 Unit.
Multilevel data analysis examples using R. Topics include: two-level nested data, growth curve modeling, generalized linear models for counts and categorical data, nonlinear models, three-level analyses. For more information, see course website: http://rogosateaching.com/stat196/.
Same as: EDUC 401D

STATS 199. Independent Study. 1-15 Unit.
For undergraduates.

STATS 200. Introduction to Statistical Inference. 3 Units.
Modern statistical concepts and procedures derived from a mathematical framework. Statistical inference, decision theory, point and interval estimation, tests of hypotheses; Neyman-Pearson theory. Bayesian analysis; maximum likelihood, large sample theory. Prerequisite: 116.

STATS 201. Design and Analysis of Experiments. 3-5 Units.
Theory and applications. Factors that affect response. Optimum levels of parameters. How to balance theory and practical design techniques. Prerequisites: basic statistics and probability theory.

STATS 202. Data Mining and Analysis. 3 Units.
Data mining is used to discover patterns and relationships in data. Emphasis is on large complex data sets such as those in very large databases or through web mining. Topics: decision trees, association rules, clustering, case based methods, and data visualization. Prereqs: Introductory courses in statistics or probability (e.g., Stats 60), linear algebra (e.g., Math 51), and computer programming (e.g., CS 105).

STATS 203. Introduction to Regression Models and Analysis of Variance. 3 Units.

STATS 204. Sampling. 3 Units.
How best to take data and where to sample it. Examples include surveys and sampling from data warehouses. Emphasis is on methods for finite populations. Topics: simple random sampling, stratified sampling, cluster sampling, ratio and regression estimators, two stage sampling.

STATS 205. Introduction to Nonparametric Statistics. 3 Units.
Nonparametric analogs of the one- and two-sample t-tests and analysis of variance; the sign test, median test, Wilcoxon's tests, and the Kruskal-Wallis and Friedman tests, tests of independence. Nonparametric regression and nonparametric density estimation, modern nonparametric techniques, nonparametric confidence interval estimates.

STATS 206. Introduction to Regression Models and Analysis of Variance. 3 Units.
Introduction to the statistical analysis of several quantitative measurements on each observational unit. Emphasis is on concepts, computer-intensive methods. Examples from economics, education, geology, psychology. Topics: multiple regression, multivariate analysis of variance, principal components, factor analysis, canonical correlations, multidimensional scaling, clustering. Pre- or corequisite: 200.

STATS 207. Introduction to Time Series Analysis. 3 Units.
Time series models used in economics and engineering. Trend fitting, autoregressive and moving average models and spectral analysis, Kalman filtering, and state-space models. Seasonality, transformations, and introduction to financial time series. Prerequisite: basic course in Statistics at the level of 200.

STATS 208. Introduction to the Bootstrap. 3 Units.
The bootstrap is a computer-based method for assigning measures of accuracy to statistical estimates. By substituting computation in place of mathematical formulas, it permits the statistical analysis of complicated estimators. Topics: nonparametric assessment of standard errors, biases, and confidence intervals; related resampling methods including the jackknife, cross-validation, and permutation tests. Theory and applications. Prerequisite: course in statistics or probability.
STATS 209. Statistical Methods for Group Comparisons and Causal Inference. 3 Units.
Critical examination of statistical methods in social science and life sciences applications, especially for cause and effect determinations. Topics: mediating and moderating variables, potential outcomes framework, encouragement designs, multilevel models, matching and propensity score methods, analysis of covariance, instrumental variables, compliance, path analysis and graphical models, group comparisons with longitudinal data. See http://rogosateaching.com/stat209/. Prerequisite: intermediate-level statistical methods. Same as: EDUC 260A, HRP 239

STATS 211. Meta-research: Appraising Research Findings, Bias, and Meta-analysis. 3 Units.
Open to graduate, medical, and undergraduate students. Appraisal of the quality and credibility of research findings; evaluation of sources of bias. Meta-analysis as a quantitative (statistical) method for combining results of independent studies. Examples from medicine, epidemiology, genomics, ecology, social/behavioral sciences, education. Collaborative analyses. Project involving generation of a meta-research project or reworking and evaluation of an existing published meta-analysis. Prerequisite: knowledge of basic statistics. Same as: CHPR 206, HRP 206, MED 206

STATS 212. Applied Statistics with SAS. 3 Units.
Data analysis and implementation of statistical tools in SAS. Topics: reading in and describing data, categorical data, dates and longitudinal data, correlation and regression, nonparametric comparisons, ANOVA, multiple regression, multivariate data analysis, using arrays and macros in SAS. Prerequisite: statistical techniques at the level of STATS 191 or 203; knowledge of SAS not required.

STATS 213. Introduction to Graphical Models. 3 Units.
Multivariate Normal Distribution and Inference, Wishart distributions, graph theory, probabilistic Markov models, pairwise and global Markov property, decomposable graph, Markov equivalence, MLE for DAG models and undirected graphical models, Bayesian inference for DAG models and undirected graphical models. Prerequisites: STATS 217, STATS 200 (preferably STATS 300A), MATH 104 or equivalent class in linear algebra. Same as: STATS 313

STATS 215. Statistical Models in Biology. 3 Units.
Poisson and renewal processes, Markov chains in discrete and continuous time, branching processes, diffusion. Applications to models of nucleotide evolution, recombination, the Wright-Fisher process, coalescence, genetic mapping, sequence analysis. Theoretical material approximately the same as in STATS 217, but emphasis is on material drawn from applications in biology, especially genetics. Prerequisite: 116 or equivalent.

STATS 216. Introduction to Statistical Learning. 3 Units.
Overview of supervised learning, with a focus on regression and classification methods. Syllabus includes: linear and polynomial regression, logistic regression and linear discriminant analysis, cross-validation and the bootstrap, model selection and regularization methods (ridge and lasso); nonlinear models, splines and generalized additive models; tree-based methods, random forests and boosting; support-vector machines; Some unsupervised learning: principal components and clustering (k-means and hierarchical). Computing is done in R, through tutorial sessions and homework assignments. This math-light course is offered via video segments (MOOC style), and in-class problem solving sessions. Prerequisite: Introductory courses in statistics or probability (e.g., Stats 60), linear algebra (e.g., Math 51), and computer programming (e.g., CS 105).

STATS 216V. Introduction to Statistical Learning. 3 Units.
Overview of supervised learning, with a focus on regression and classification methods. Syllabus includes: linear and polynomial regression, logistic regression and linear discriminant analysis, cross-validation and the bootstrap, model selection and regularization methods (ridge and lasso); nonlinear models, splines and generalized additive models; tree-based methods, random forests and boosting; support-vector machines; Some unsupervised learning: principal components and clustering (k-means and hierarchical). Computing is done in R, through tutorial sessions and homework assignments. This math-light course is offered remotely only via video segments (MOOC style). TAs will host remote weekly office hours using an online platform such as Google Hangout or BlueJeans. There are four homework assignments, a midterm, and a final exam, all of which are administered remotely. Prerequisites: introductory courses in statistics or probability (e.g., Stats 60), linear algebra (e.g., Math 51), and computer programming (e.g., CS 105).

STATS 217. Introduction to Stochastic Processes I. 2-3 Units.
Discrete and continuous time Markov chains, poisson processes, random walks, branching processes, first passage times, recurrence and transience, stationary distributions. Non-Statistics masters students may want to consider taking STATS 215 instead. Prerequisite: STATS 116 or consent of instructor.

STATS 218. Introduction to Stochastic Processes II. 3-4 Units.
Renewal theory, Brownian motion, Gaussian processes, second order processes, martingales.

STATS 219. Stochastic Processes. 3 Units.

STATS 221. Introduction to Mathematical Finance. 3-4 Units.

STATS 222. Statistical Methods for Longitudinal Research. 2-3 Units.
Research designs and statistical procedures for time-ordered (repeated-measures) data. The analysis of longitudinal panel data is central to empirical research on learning, development, aging, and the effects of interventions. Topics include: measurement of change, growth curve models, analysis of durations including survival analysis, experimental and non-experimental group comparisons, reciprocal effects, stability. See http://rogosateaching.com/stat222/. Prerequisite: intermediate statistical methods. Same as: EDUC 351A

STATS 229. Machine Learning. 3-4 Units.
Topics: statistical pattern recognition, linear and non-linear regression, non-parametric methods, exponential family, GLMs, support vector machines, kernel methods, model/feature selection, learning theory, VC dimension, clustering, density estimation, EM, dimensionality reduction, ICA, PCA, reinforcement learning and adaptive control, Markov decision processes, approximate dynamic programming, and policy search. Prerequisites: linear algebra, and basic probability and statistics. Same as: CS 229
STATS 231. Statistical Learning Theory. 3 Units.
How do we formalize what it means for an algorithm to learn from data?
This course focuses on developing mathematical tools for answering this question. We will present various common learning algorithms and prove theoretical guarantees about them. Topics include classical asymptotics, method of moments, generalization bounds via uniform convergence, kernel methods, online learning, and multi-armed bandits. Prerequisites: A solid background in linear algebra and probability theory, statistics and machine learning (STATS 315A or CS 229). Convex optimization (EE 364A) is helpful but not required.
Same as: CS 229T

STATS 237. Theory of Investment Portfolios and Derivative Securities. 3 Units.

STATS 237P. Theory of Investment Portfolios and Derivative Securities. 3 Units.
For SCPD students; see STATS237.

STATS 238. The Future of Finance. 2 Units.
If you are interested in a career in finance or that touches finance (computational science, economics, public policy, legal, regulatory, corporate, other), this course will give you a useful perspective. We will take on hot topics in the current landscape of the global markets as the world continues to evolve from the financial crisis. We will discuss the sweeping change underway at the policy level by regulators and legislators around the world and how this is changing business models for existing players and attracting new players to finance. The course will include guest-lecturer perspectives on where the greatest opportunities exist for students entering or touching the world of finance today including new and disruptive players in fin tech, crowd financing, block chain, robo advising, algorithmic trading, big data and other areas.
New challenges such as cyber and financial warfare threats also will be addressed. While derivatives and other quantitative concepts will be handled in a non-technical way, some knowledge of finance and the capital markets is presumed. Elements used in grading: Class Participation, Attendance, Final Paper. Consent Application: To apply for this course, students must complete and email to the instructors the Consent Application Form, which is available on the Public Policy Program’s website at https://publicpolicy.stanford.edu/academics/undergraduate/forms. See Consent Application Form for submission deadline. (Cross-listed as ECON252/152, PUBLPOL364, STATS238, LAW 1038.).
Same as: ECON 152, ECON 252, PUBLPOL 364

STATS 239. Mathematical and Computational Finance Seminar. 1 Unit.
Same as: CME 242, MS&E 446A

STATS 239A. Workshop in Quantitative Finance. 1 Unit.
Topics of current interest.

STATS 239B. Workshop in Quantitative Finance. 1 Unit.
Topics of current interest. May be repeated for credit.
Same as: CME 239B

STATS 240. Statistical Methods in Finance. 3-4 Units.

STATS 240P. Statistical Methods in Finance. 3 Units.
For SCPD students; see 240.

STATS 241. Data-driven Financial and Risk Econometrics. 3-4 Units.

STATS 241P. Data-driven Financial and Risk Econometrics. 3 Units.
For SCPD students; see STATS241.

STATS 243. Risk Analytics and Management in Finance and Insurance. 3 Units.
Same as: CME 243

STATS 243P. Risk Analytics and Management in Finance and Insurance. 3 Units.
For SCPD students; see STATS243.

STATS 244. Quantitative Trading: Algorithms, Data, and Optimization. 2-4 Units.

STATS 244P. Quantitative Trading: Algorithms, Data, and Optimization. 3 Units.
For SCPD students; see 244.

STATS 245. Data, Models, and Decision Analytics. 3 Units.
Statistical models and decision theory. Online A/B testing, comparative effective studies of medical treatments. Introduction to recommender systems in online services, personalized medicine and marketing. Prerequisite or corequisite: STATS 202, or CS 229, or CME 250, or equivalent.

STATS 245P. Data, Models, and Decision Analytics. 3 Units.
For SCPD students; see STATS245.

STATS 247. Data-driven Decisions in the Big Data Era of Economics and Technology. 3 Units.
(SCPD students register for 247P). Description TBD.

STATS 247P. Data-driven Decisions in the Big Data Era of Economics and Technology. 3 Units.
For SCPD students; see STATS247.

STATS 250. Mathematical Finance. 3 Units.
Same as: MATH 238

STATS 253. Analysis of Spatial and Temporal Data. 3 Units.
A unified treatment of methods for spatial data, time series, and other correlated data from the perspective of regression with correlated errors. Two main paradigms for dealing with autocorrelation: covariance modeling (kriging) and autoregressive processes. Bayesian methods. Prerequisites: applied linear algebra (MATH 103 or equivalent), statistical estimation (STATS 200 or CS 229), and linear regression (STATS 203 or equivalent).
STATS 260A. Workshop in Biostatistics. 1-2 Unit.
Applications of statistical techniques to current problems in medical science. To receive credit for one or two units, a student must attend every workshop. To receive two units, in addition to attending every workshop, the student is required to write an acceptable one-page summary of two of the workshops, with choices made by the student. Same as: BIODS 260A

STATS 260B. Workshop in Biostatistics. 1-2 Unit.
Applications of statistical techniques to current problems in medical science. To receive credit for one or two units, a student must attend every workshop. To receive two units, in addition to attending every workshop, the student is required to write an acceptable one-page summary of two of the workshops, with choices made by the student. Same as: BIODS 260B

STATS 260C. Workshop in Biostatistics. 1-2 Unit.
Applications of statistical techniques to current problems in medical science. To receive credit for one or two units, a student must attend every workshop. To receive two units, in addition to attending every workshop, the student is required to write an acceptable one-page summary of two of the workshops, with choices made by the student. Same as: BIODS 260C

STATS 261. Intermediate Biostatistics: Analysis of Discrete Data. 3 Units.
Methods for analyzing data from case-control and cross-sectional studies: the 2x2 table, chi-square test, Fisher’s exact test, odds ratios, Mantel-Haenszel methods, stratification, tests for matched data, logistic regression, conditional logistic regression. Emphasis is on data analysis in SAS. Special topics: cross-fold validation and bootstrap inference. Same as: BIOMEDIN 233, HRP 261

STATS 262. Intermediate Biostatistics: Regression, Prediction, Survival Analysis. 3 Units.
Methods for analyzing longitudinal data. Topics include Kaplan-Meier methods, Cox regression, hazard ratios, time-dependent variables, longitudinal data structures, profile plots, missing data, modeling change, MANOVA, repeated-measures ANOVA, GEE, and mixed models. Emphasis is on practical applications. Prerequisites: basic ANOVA and linear regression. Same as: HRP 262

STATS 263. Design of Experiments. 3 Units.

STATS 266. Advanced Statistical Methods for Observational Studies. 2-3 Units.
Design principles and statistical methods for observational studies. Topics include: matching methods, sensitivity analysis, instrumental variables, graphical models, marginal structural models. 3 unit registration requires a small project and presentation. Computing is in R. Pre-requisites: HRP 261 and 262 or STAT 209 (HRP 239), or equivalent. See http://rogosateaching.com/somgen290/.
Same as: CHPR 290, EDUC 260B

STATS 267. Probability: Ten Great Ideas About Chance. 4 Units.
Foundational approaches to thinking about chance in matters such as gambling, the law, and everyday affairs. Topics include: chance and decisions; the mathematics of chance; frequencies, symmetry, and chance; Bayes great idea; chance and psychology; measures of chance; and harnessing chance. Emphasis is on the philosophical underpinnings and problems. Prerequisite: exposure to probability or a first course in statistics at the level of STATS 60 or 116. Same as: PHIL 166, PHIL 266, STATS 167

STATS 270. Bayesian Statistics I. 3 Units.
This is the first of a two course sequence on modern Bayesian statistics. Topics covered include: real world examples of large scale Bayesian analysis; basic tools (models, conjugate priors and their mixtures); Bayesian estimates, tests and credible intervals; foundations (axioms, exchangeability, likelihood principle); Bayesian computations (Gibbs sampler, data augmentation, etc.); prior specification. Prerequisites: statistics and probability at the level of Stats300A, Stats305, and Stats310. Same as: STATS 370

STATS 271. Bayesian Statistics II. 3 Units.
This is the second of a two course sequence on modern Bayesian statistics. Topics covered include: Asymptotic properties of Bayesian procedures and consistency (Doobs theorem, frequentists consistency, counter examples); connections between Bayesian methods and classical methods (the complete class theorem); generalization of exchangeability; general versions of the Bayes theorem in the undominated case; non parametric Bayesian methods (Dirichlet and Polya tree priors). Throughout general theory will be illustrated with classical examples. Prerequisites: Stats 270/370.
Same as: STATS 371

STATS 279. Paradigms for Computing with Data. 3 Units.
Advanced programming and computing techniques to support projects in data analysis and related research. For Statistics graduate students and others whose research involves data analysis and development of associated computational software. Prerequisites: Programming experience including familiarity with R; computing at least at the level of CS 106, statistics at the level of STATS 110 or 141.

STATS 289. Industrial Research for Statisticians. 1 Unit.
Masters-level research as in 299, but with the approval and supervision of a faculty adviser, it must be conducted for an off-campus employer. Students must submit a written final report upon completion of the internship in order to receive credit. Repeatable for credit. Prerequisite: enrollment in Statistics M.S. program.

STATS 299. Independent Study. 1-10 Unit.
For Statistics M.S. students only. Reading or research program under the supervision of a Statistics faculty member. May be repeated for credit.

Main topic: statistical inference of latent variable models (including SBM), using EM-like algorithms. The critical step is the determination of the conditional distribution of the latent variables given the observed data, which is doable for mixture models and hidden Markov models. For more complex models such as the stochastic block model (SBM: popular in sociology, physics, biology, etc.) variational approximations can be used to derive a generalized version of EM algorithm. This approach can be extended to Bayesian inference (variational Bayes EM algorithm). If time permits, change-point detection models will be introduced. Topics will be illustrated with examples from genomics.

STATS 300A. Theory of Statistics I. 2-3 Units.
Finite sample optimality of statistical procedures; Decision theory: loss, risk, admissibility; Principles of data reduction: sufficiency, ancillarity, completeness; Statistical models: exponential families, group families, nonparametric families; Point estimation: optimal unbiased and equivariant estimation, Bayes estimation, minimax estimation; Hypothesis testing and confidence intervals: uniformly most powerful tests, uniformly most accurate confidence intervals, optimal unbiased and invariant tests. Prerequisites: Real analysis, introductory probability (at the level of STATS 116), and introductory statistics.
STATS 300B. Theory of Statistics II. 2-4 Units.
Elementary decision theory; loss and risk functions, Bayes estimation; UMVU estimator, minimax estimators, shrinkage estimators. Hypothesis testing and confidence intervals: Neyman-Pearson theory; UMP tests and uniformly most accurate confidence intervals; use of unbiasedness and invariance to eliminate nuisance parameters. Large sample theory: basic convergence concepts; robustness; efficiency; contiguity, locally asymptotically normal experiments; convolution theorem; asymptotically UMP and maximin tests. Asymptotic theory of likelihood ratio and score tests. Rank permutation and randomization tests; jackknife, bootstrap, subsampling and other resampling methods. Further topics: sequential analysis, optimal experimental design, empirical processes with applications to statistics, Edgeworth expansions, density estimation, time series.

STATS 300C. Theory of Statistics III. 2-4 Units.
Decision theory formulation of statistical problems. Minimax, admissible procedures. Complete class theorems ("all minmax or admissible procedures are "Bayes"), Bayes procedures, conjugate priors, hierarchical models. Bayesian non parametrics: diaichlet, tail free, polya trees, bayesian sieves. Inconsistency of bayes rules.

STATS 302. Qualifying Exams Workshop. 3 Units.
Prepares Statistics Ph.D. students for the qualifying exams by reviewing relevant course topics and problem solving strategies.

STATS 303. PhD First Year Student Workshop. 1 Unit.
For Statistics First Year PhD students only. Discussion of relevant topics in first year student courses, consultation with PhD advisor.

STATS 305A. Introduction to Statistical Modeling. 3 Units.

STATS 305B. Methods for Applied Statistics I. 3 Units.
Regression modeling extended to categorical data. Logistic regression. Loglinear models. Generalized linear models. Discriminant analysis. Categorical data models from information retrieval and Internet modeling. Prerequisite: 305A or equivalent. (NB: prior to 2016-17 the 305ABC series was numbered as 305, 306A and 306B).

STATS 305C. Methods for Applied Statistics II: Applied Bayesian Statistics. 3 Units.
Applied Bayesian statistics. Fundamentals, hierarchical models, computing. (NB: prior to 2016-17 the 305ABC series was numbered as 305, 306A and 306B).

STATS 310A. Theory of Probability I. 2-4 Units.
Mathematical tools: sigma algebras, measure theory, connections between coin tossing and Lebesgue measure, basic convergence theorems. Probability: independence, Borel-Cantelli lemmas, almost sure and Lp convergence, weak and strong laws of large numbers. Large deviations. Weak convergence; central limit theorems; Poisson convergence; Stein’s method. Prerequisites: 116, MATH 171. Same as: MATH 230A

STATS 310B. Theory of Probability II. 2-3 Units.
Conditional expectations, discrete time martingales, stopping times, uniform integrability, applications to 0-1 laws, Radon-Nikodym Theorem, ruin problems, etc. Other topics as time allows selected from (i) local limit theorems, (ii) renewal theory, (iii) discrete time Markov chains, (iv) random walk theory, (v) ergodic theory. Prerequisite: 310A or MATH 230A.

Same as: MATH 230B

STATS 310C. Theory of Probability III. 2-4 Units.

Same as: MATH 230C

STATS 311. Information Theory and Statistics. 3 Units.
Information theoretic techniques in probability and statistics. Fano, Assouad, Le Cam methods for optimality guarantees in estimation. Large deviations and concentration inequalities (Sanov’s theorem, hypothesis testing, thentropy method, concentration of measure). Approximation of (Bayes) optimal procedures, surrogate risks, f-divergences. Penalized estimators and minimumdescription length. Online game playing, gambling, no-regret learning. Prerequisites: EE 376A (or equivalent) or STATS 300A.

Same as: EE 377

STATS 312. Statistical Methods in Neuroscience. 3 Units.
The goal is to discuss statistical methods for neuroscience in their natural habitat: the research questions, measurement technologies and experiment designs used in modern neuroscience. We will emphasize both the choice and quality of the methods, as well as the reporting, interpretation and visualization of results. Likely topics include preprocessing and signal extraction for single-neuron and neuroimaging technologies, statistical models for single response, encoding and decoding models, multiple-responses and parametric maps, and testing. Participation includes analyzing methods and real data, discussing papers in class, and a final project. Requirements: we will assume familiarity with linear models, likelihoods etc. Students who have not taken graduate level statistics courses are required to contact the instructor. Background in neuroscience is not assumed.

STATS 313. Introduction to Graphical Models. 3 Units.
Multivariate Normal Distribution and Inference, Wishart distributions, graph theory, probabilistic Markov models, pairwise and global Markov property, decomposable graph, Markov equivalence, MLE for DAG models and undirected graphical models, Bayesian inference for DAG models and undirected graphical models. Prerequisites: STATS 217, STATS 200 (preferably STATS 300A), MATH 104 or equivalent class in linear algebra. Same as: STATS 213

STATS 314A. Advanced Statistical Theory. 3 Units.
Covers a range of topics, including: empirical processes, asymptotic efficiency, uniform convergence of measures, contiguity, resampling methods, Edgeworth expansions.

STATS 314B. Topics in Minimax Inference of Nonparametric Functionals. 3 Units.
Topics in the estimation of various functionals of underlying distribution for nonparametric problems. Development of ideas of higher order influence functions that extend the theory of classical first order semiparametric theory. Topics on adaptive estimation and adaptive confidence sets construction. Understanding results from wavelet theory and higher order U-statistics.
STATS 315A. Modern Applied Statistics: Learning. 2-3 Units.

STATS 315B. Modern Applied Statistics: Data Mining. 2-3 Units.
Two-part sequence. New techniques for predictive and descriptive learning using ideas that bridge gaps among statistics, computer science, and artificial intelligence. Emphasis is on statistical aspects of their application and integration with more standard statistical methodology. Predictive learning refers to estimating models from data with the goal of predicting future outcomes, in particular, regression and classification models. Descriptive learning is used to discover general patterns and relationships in data without a predictive goal, viewed from a statistical perspective as computer automated exploratory analysis of large complex data sets.

STATS 316. Stochastic Processes on Graphs. 1-3 Unit.
Local weak convergence, Gibbs measures on trees, cavity method, and replica symmetry breaking. Examples include random k-satisfiability, the assignment problem, spin glasses, and neural networks. Prerequisite: 310A or equivalent.

STATS 317. Stochastic Processes. 3 Units.

STATS 318. Modern Markov Chains. 3 Units.
Tools for understanding Markov chains as they arise in applications. Random walk on graphs, reversible Markov chains, Metropolis algorithm, Gibbs sampler, hybrid Monte Carlo, auxiliary variables, hit and run, Swedson-Wong algorithms, geometric theory, Poincare-Nash-Cheger-Log-Sobolov inequalities. Comparison techniques, coupling, stationary times, Harris recurrence, central limit theorems, and large deviations.

STATS 319. Literature of Statistics. 1-3 Unit.
Literature study of topics in statistics and probability culminating in oral and written reports. May be repeated for credit.

STATS 320. Heterogeneous Data with Kernels. 3 Units.


STATS 322. Function Estimation in White Noise. 2-3 Units.

STATS 324. Multivariate Analysis. 2-3 Units.
Classic multivariate statistics: properties of the multivariate normal distribution, determinants, volumes, projections, matrix square roots, the singular value decomposition; Wishart distributions, Hotelling's T-square; principal components, canonical correlations, Fisher's discriminant, the Cauchy projection formula.

STATS 325. Multivariate Analysis and Random Matrices in Statistics. 2-3 Units.
Topics on Multivariate Analysis and Random Matrices in Statistics (full description TBA).

STATS 329. Large-Scale Simultaneous Inference. 1-3 Unit.
Estimation, testing, and prediction for microarray-like data. Modern scientific technologies, typified by microarrays and imaging devices, produce inference problems with thousands of parallel cases to consider simultaneously. Topics: empirical Bayes techniques, James-Stein estimation, large-scale simultaneous testing, false discovery rates, local fdr, proper choice of null hypothesis (theoretical, permutation, empirical nulls), power, effects of correlation on tests and estimation accuracy, prediction methods, related sets of cases ("enrichment"), effect size estimation. Theory and methods illustrated on a variety of large-scale data sets.

STATS 330. An Introduction to Compressed Sensing. 3 Units.
Compressed sensing is a new data acquisition theory asserting that one can design nonadaptive sampling techniques that condense the information in a compressible signal into a small amount of data. This revelation may change the way engineers think about signal acquisition. Course covers fundamental theoretical ideas, numerical methods in large-scale convex optimization, hardware implementations, connections with statistical estimation in high dimensions, and extensions such as recovery of data matrices from few entries (Famous Netflix Prize). Same as: CME 362

STATS 331. Survival Analysis. 2 Units.
The course introduces basic concepts, theoretical basis and statistical methods associated with survival data. Topics include censoring, Kaplan-Meier estimation, logrank test, proportional hazards regression, accelerated failure time model, multivariate failure time analysis and competing risks. The traditional counting process/martingale methods as well as modern empirical process methods will be covered. Prerequisite: Understanding of basic probability theory and statistical inference methods. Same as: BIODS 231
STATS 333. Modern Spectral Analysis. 3 Units.
Traditional spectral analysis encompassed Fourier methods and their elaborations, under the assumption of a simple superposition of sinusoids, independent of time. This enables development of efficient and effective computational schemes, such as the FFT. Since many systems change in time, it becomes of interest to generalize classical spectral analysis to the time-varying setting. In addition, classical methods suffer from resolution limits which we hope to surpass. In this topics course, we follow two threads. On the one hand, we consider the estimation of instantaneous frequencies and decomposition of source signals, which may be time-varying. The thread begins with the empirical mode decomposition (EMD) for non-stationary signal decomposition into intrinsic mode functions (IMFs), introduced by N. Huang et al [1], together with its machinery of the sifting process and computation of the Hilbert spectrum, resulting in the so-called adaptive harmonic model (AHM). Next, this thread considers the wavelet synchrosqueezing transform (WSST) proposed by Daubechies et al [2], which attempts to estimate instantaneous frequencies (IFs), via the frequency re-assignment (FRA) rule, that facilitates non-stationary signal decomposition. In reference [3], a real-time method is proposed for computing the FRA rule; and in reference [4], the exact number of AHM components is determined with more precise estimation of the IFs, for more accurate extraction of the signal components and polynomial-like trend. On another thread, recent developments in optimization have been applied to obtain time-varying spectra or very high-resolution spectra; in particular, references [5]-[8] give examples of recent results where convex estimation is applied to obtain new and more highly resolved spectral estimates, some with time-varying structure.

STATS 338. Topics in Biostatistics. 3 Units.

STATS 341. Applied Multivariate Statistics. 3 Units.
Theory, computational aspects, and practice of a variety of important multivariate statistical tools for data analysis. Topics include classical multivariate Gaussian and undirected graphical models, graphical displays, PCA, SVD and generalizations including canonical correlation analysis, linear discriminant analysis, correspondence analysis, with focus on recent variants. Factor analysis and independent component analysis. Multidimensional scaling and its variants (e.g. Isomap, spectral clustering). Students are expected to program in R. Prerequisite: STATS 305 or equivalent.

STATS 344. Introduction to Statistical Genetics. 3 Units.
Statistical methods for analyzing human genetics studies of Mendelian disorders and common complex traits. Probable topics include: principles of population genetics; epidemiologic designs; familial aggregation; segregation analysis; linkage analysis; linkage-disequilibrium-based association mapping approaches; and genome-wide analysis based on high-throughput genotyping platforms. Prerequisite: STATS 116 or equivalent or consent of instructor.

Same as: GENE 244

STATS 345. Statistical and Machine Learning Methods for Genomics. 3 Units.
Introduction to statistical and computational methods for genomics. Sample topics include: expectation maximization, hidden Markov model, Markov chain Monte Carlo, ensemble learning, probabilistic graphical models, kernel methods and other modern machine learning paradigms. Rationales and techniques illustrated with existing implementations used in population genetics, disease association, and functional regulatory genomics studies. Instruction includes lectures and discussion of readings from primary literature. Homework and projects require implementing some of the algorithms and using existing toolkits for analysis of genomic datasets.

Same as: BIO 268, BIOMEDIN 245, CS 373, GENE 245

Concentration of measure techniques. Mean field models for disordered systems: infinite size limit, computing the free energy, ultrametricity, dynamics. Interpolation techniques and infinite size limit in information theory and coding. May be repeated once for credit. Prerequisite: 310A or equivalent.

STATS 351. Random Walks, Networks and Environment. 3 Units.
Selected material about probability on trees and networks, random walk in random and non-random environments, percolation and related interacting particle systems. Prerequisite: Exposure to measure theoretic probability and to stochastic processes.

STATS 351A. An Introduction to Random Matrix Theory. 3 Units.
Patterns in the eigenvalue distribution of typical large matrices, which also show up in physics (energy distribution in scattering experiments), combinatorics (length of longest increasing subsequence), first passage percolation and number theory (zeros of the zeta function). Classical compact ensembles (random orthogonal matrices). The tools of determinantal point processes.

Same as: MATH 231A

STATS 355. Observational Studies. 2-3 Units.
This course will cover statistical methods for the design and analysis of observational studies. Topics for the course will include the potential outcomes framework for causal inference; randomized experiments; methods for controlling for observed confounders in observational studies; sensitivity analysis for hidden bias; instrumental variables; tests of hidden bias; coherence; and design of observational studies.

Same as: HRP 255

STATS 359. Topics in Mathematical Physics. 3 Units.
Covers a list of topics in mathematical physics. The specific topics may vary from year to year, depending on the instructor’s discretion. Background in graduate level probability theory and analysis is desirable.

Same as: MATH 273

STATS 360. Advanced Statistical Methods for Earth System Analysis. 3 Units.
Introduction for graduate students to important issues in data analysis relevant to earth system studies. Emphasis on methodology, concepts and implementation (in R), rather than formal proofs. Likely topics include the bootstrap, non-parametric methods, regression in the presence of spatial and temporal correlation, extreme value analysis, time-series analysis, high-dimensional regressions and change-point models. Topics subject to change each year. Prerequisites: STATS 110 or equivalent.

Same as: ESS 260

STATS 362. Topic: Monte Carlo. 3 Units.

STATS 363. Design of Experiments. 3 Units.

Same as: STATS 263
STATS 366. Modern Statistics for Modern Biology. 3 Units.
Application based course in nonparametric statistics. Modern toolbox of visualization and statistical methods for the analysis of data, examples drawn from immunology, microbiology, cancer research and ecology. Methods covered include multivariate methods (PCA and extensions), sparse representations (trees, networks, contingency tables) as well as nonparametric testing (Bootstrap, permutation and Monte Carlo methods). Hands on, use R and cover many Bioconductor packages. Prerequisite: Minimal familiarity with computers. Instructor consent. Location: Li Ka Shing Center, room 120. Same as: BIOS 221

STATS 367. Statistical Models in Genetics. 3 Units.
This course will cover statistical problems in population genetics and molecular evolution with an emphasis on coalescent theory. Special attention will be paid to current research topics, illustrating the challenges presented by genomic data obtained via high-throughput technologies. No prior knowledge of genomics is necessary. Familiarity with the R statistical package or other computing language is needed for homework assignments. Prerequisites: knowledge of probability through elementary stochastic processes and statistics through likelihood theory.

STATS 368. Empirical Process Theory and its Applications. 3 Units.
This course is on the theory of empirical processes. In the course we will focus on weak convergence of stochastic processes, M-estimation and empirical risk minimization. The course will cover topics like covering numbers and bracketing numbers, maximal inequalities, chaining and symmetrization, uniform law of large numbers and uniform central limit theorems, rates of convergence of MLEs and (penalized) least squares estimators, and concentration inequalities.

STATS 369. Methods from Statistical Physics. 3 Units.
Mathematical techniques from statistical physics have been applied with increasing success on problems form combinatorics, computer science, machine learning. These methods are non-rigorous, but in several cases they were proved to yield correct predictions. This course provides a working knowledge of these methods for non-physicists. Specific topics: the Sherrington-Kirkpatrick model; sparse regression with random designs.

STATS 370. Bayesian Statistics I. 3 Units.
This is the first of a two course sequence on modern Bayesian statistics. Topics covered include: real world examples of large scale Bayesian analysis; basic tools (models, conjugate priors and their mixtures); Bayesian estimates, tests and credible intervals; foundations (axioms, exchangeability, likelihood principle); Bayesian computations (Gibbs sampler, data augmentation, etc.); prior specification. Prerequisites: statistics and probability at the level of Stats300A, Stats305, and Stats310.
Same as: STATS 270

STATS 371. Bayesian Statistics II. 3 Units.
This is the second of a two course sequence on modern Bayesian statistics. Topics covered include: Asymptotic properties of Bayesian procedures and consistency (Doobs theorem, frequentists consistency, counter examples); connections between Bayesian methods and classical methods (the complete class theorem); generalization of exchangeability; general versions of the Bayes theorem in the undominated case; non parametric Bayesian methods (Dirichelet and Polya tree priors). Throughout general theory will be illustrated with classical examples. Prerequisites: Stats270/370.
Same as: STATS 271

STATS 374. Large Deviations Theory. 3 Units.
Same as: MATH 234

STATS 375. Inference in Graphical Models. 3 Units.
Graphical models as a unifying framework for describing the statistical relationships between large sets of variables; computing the marginal distribution of one or a few such variables. Focus is on sparse graphical structures, low-complexity algorithms, and their analysis. Topics include: variational inference; message passing algorithms; belief propagation; generalized belief propagation; survey propagation. Analysis techniques: correlation decay; distributional recursions. Applications from engineering, computer science, and statistics. Prerequisite: EE 278, STATS 116, or CS 228. Recommended: EE 376A or STATS 217.

STATS 376A. Information Theory. 3 Units.
The fundamental ideas of information theory. Entropy and intrinsic randomness. Data compression to the entropy limit. Huffman coding. Arithmetic coding. Channel capacity, the communication limit. Gaussian channels. Kolmogorov complexity. Asymptotic equipartition property. Information theory and Kelly gambling. Applications to communication and data compression. Prerequisite: EE178 or STATS 116, or equivalent. Same as: EE 376A

STATS 376B. Network Information Theory. 3 Units.
Network information theory deals with the fundamental limits on information flow in networks and the optimal coding schemes that achieve these limits. It aims to extend Shannon's point-to-point information theory and the Ford-Fulkerson max-flow min-cut theorem to networks with multiple sources and destinations. The course presents the basic results and tools in the field in a simple and unified manner. Topics covered include: multiple access channels, broadcast channels, interference channels, channels with state, distributed source coding, multiple description coding, network coding, relay channels, interactive communication, and noisy network coding. Prerequisites: EE376A.
Same as: EE 376B

STATS 390. Consulting Workshop. 1-3 Unit.
Skills required of practicing statistical consultants, including exposure to statistical applications. Students participate as consultants in the department's drop-in consulting service, analyze client data, and prepare formal written reports. Seminar provides supervised experience in short term consulting. May be repeated for credit. Prerequisites: course work in applied statistics or data analysis, and consent of instructor.

STATS 396. Research Workshop in Computational Biology. 1-2 Unit.
Applications of Computational Statistics and Data Mining to Biological Data. Attendance mandatory. Instructor approval required.

STATS 397. PhD Oral Exam Workshop. 1 Unit.
For Statistics PhD students defending their dissertation.

STATS 398. Industrial Research for Statisticians. 1 Unit.
Data. Attendance mandatory. Instructor approval required.

STATS 399. Research. 1-10 Unit.
Research work as distinguished from independent study of nonresearch character listed in 199. May be repeated for credit.

STATS 801. TGR Project. 0 Units.

STATS 802. TGR Dissertation. 0 Units.