III. THE PH.D. PROGRAM

The Doctor of Philosophy degree program with a major in statistics is designed for students who plan to pursue careers in university teaching and research or in industrial and government research and consulting. A doctoral student pursuing the degree program in statistics may choose to emphasize either statistics or probability.

In addition to meeting the requirements of the university and of the College of Natural Science, students must meet the requirements specified below.

Admission

A master's level understanding of statistics and probability and a sound understanding of undergraduate-level real analysis are necessary for success in the doctoral program. Strong applicants with deficiencies in one of these areas will be considered for admission, and if accepted will be given the opportunity to learn the required material during their first year in the program. The Graduate Record Examination (GRE) General Test is required for all applicants.

III.1 The Guidance Committee

The Graduate Director serves as the academic advisor for every student as soon as he/she enters the Ph.D. program. The advisor must approve the course selection of the student.

The Guidance Committee will be formed when the student passes both prelims and selects a thesis advisor. This normally will occur during the second year. The committee will consist of at least three faculty members of this Department and at least one from outside the Department as determined by the thesis advisor in consultation with the student. The primary responsibility of the committee will be to advise the student in his/her thesis research.

III.2 Course Work

STT 872, STT 881-2 and STT 867-868 are the core courses. In addition: 8 courses as follows:

1. At least 5 courses from among (a) and (b):

    (a) Advanced Probability: STT 961, 962, 964, 996 (at least 1)

    (b) Advanced Statistics: STT 873, 874, 951, 953, 997 (at least 1)

2. At least 3 electives: Graduate courses taken inside or outside the Department.
NOTE: STT 996 and STT 997 are special topic courses, which may change from year to year. Descriptions of courses can be found at: https://reg.msu.edu/Courses/Search.aspx.

A working knowledge of real analysis is required for successful completion of the Ph.D. program. Students without sufficient background must take a course in analysis, e.g. MTH 828.

Contents of STT 867, 868, 872, 873, 874, 953, 954, 961, 964, 996, 997 are described later. These courses and the semesters they are offered is as follows:

867	Linear Models Methodology	(Fall, every year)
868	Mixed Models: Theory, Methods, and Applications	(Spring, every year)
872	Statistical Inference I	(Spring, every year)
873	Statistical Learning and Data Mining	(Fall, odd years)
874	Introduction to Bayesian Analysis	(Fall, even years)
951	Statistical Inference II	(Spring, odd years)
953	Asymptotic theory	(Spring, even years)
961	Weak convergence and Asymptotic Theory	(Fall, odd years)
962	Fractional Processes and Power Laws	(Spring, even years)
964	Stochastic Analysis	(Spring, even years)
996	Advanced Topics in Probability	(Fall, Spring, Summer)
997	Advanced Topics in Statistics	(Fall, Spring, Summer)
999	Phd Dissertation Research Credits	24 minimum; 36 maximum

Ph.D. Program (Year 1)

At the time a student arrives, he/she will consult with the Graduate Director regarding the student's level of preparation. Students with sufficient understanding of probability and statistics at the master's level and a rigorous course on real analysis as represent by MTH 828 at MSU will be expected to take courses listed under the year 1 program below. In some cases, a student may be asked to enroll in a combination of courses. For example, a student with a strong background in real analysis, but relatively little in statistics might take STT 881-2, STT 861-862, 802 and an elective during his/her first year, but not STT 867-8 or 872.

Program I: Doctoral Courses for Year 1

Fall	Sem. Cr.	**Spring**	Sem. Cr.
STT 867	3	STT 868	3
STT 881	3	STT 882	3
Elective	3	STT 872	3

NOTE: STT 881-2 is a rigorous sequence in measure theory and probability. STT 872 is a rigorous course in mathematical statistics. STT 867-8 is a sequence in theory-based statistical methodology.

Ph.D. Program (Year 2)

During the second academic year, the student will be expected to take three courses each semester. For students who take STT 861-862, 881-882 and 802 during their first year here, this would be their third year program. A student's program is always subject to the approval by the Graduate

Director or the Guidance Committee. The Program I is treated as fast track. It is advisable that the student be careful in choosing the course work.

Summer offerings

The offerings of the Department in the summers are light. Students usually use the time to prepare for examinations or for thesis research. Two graduate courses are offered each summer with topics varying, STT 890 (master's level) and either STT 996 or 997 (doctoral level). In the past few summers, graduate courses were offered in nonparametric density estimation, Bayesian statistics, analysis of contingency tables, chaos theory, stochastic models in biology, finance, inference for stationary processes, reliability and survival analysis, limit theorems for dependent random variables, curve estimation and wavelets, statistical inference for Markov and general stochastic processes, graphical methods in regression, statistical inference for images, spatial statistics, and in statistical machine learning. The senior students (2/3+) are encouraged to look for internship in research labs, government and in high profile industry.

III.3 Examinations

Prelim exams are given in the week before the start of Fall and Spring Semesters.

Prelim exams has two parts: (a) Probability prelim based on STT881-882, and (b) Statistics prelim based on STT 867, 868 and 872. In order to remain in the doctoral program, students are required to pass both exams which must be taken in the semester following their enrollment in STT 867-8, 872, 881-2. To continue in the Ph.D. program, a student must pass the exams in at most two attempts for each and within 3 years of being admitted to the Ph.D. program.

III.4 Attendance at Colloquia and Seminars

The Department has colloquia which meet regularly on Tuesdays at 10:20-11:10 a.m. Speakers are either members of the Department or are statisticians, probabilists, or those conducting research in other disciplines related to statistics or probability. The Department strongly recommends attendance by doctoral students, and occasionally, by masters students. In addition, seminars on statistics or probability are regularly scheduled in which both professors and students participate.

III.5 Thesis

A doctoral candidate must demonstrate the ability to carry out significant original research in statistics and probability. This is accomplished through the writing of a dissertation under the direction of a thesis advisor. It is the responsibility of the student to find a thesis advisor. During the spring semester of the last but one year of a student's doctoral study, the student should present a thesis plan to the Guidance Committee.

Copies of theses written by former students are available in the Department's Katz Library. The candidate must present the results of his/her thesis research in a talk of approximately one hour open to the public. After that presentation, the Guidance Committee meets further in a closed session to determine whether the student's thesis should be approved. The thesis must be submitted in an electronic form acceptable to the Graduate School. Deadline dates are available from the

Graduate School web page https://grad.msu.edu/etd/etd-deadline-dates. Two bound copies of the thesis must be submitted to the Department. The thesis is stored in electronic form by the university.

III.6 Annual Progress Report for Ph.D. Students

During each spring semester, the student must submit an "Annual Progress Report for Ph.D. Students." This form is available from https://grad.msu.edu/sites/default/files/content/forms/progressreportphd.pdf. The chairperson of the student's Guidance Committee is responsible for completion of the second page of the report. Students who hold assistantships and international students on F-1 visas are expected to complete nine credits which contribute to the completion of the degree each of Fall and Spring semesters. Those holding assistantships during the summer must complete at least three credits. **Students who have completed their coursework that is listed by their Guidance Committee can begin enrolling in STT 999. The University requires that a Ph.D. candidate complete 24 credits of 999 courses (thesis research) to graduate but no more than 36. Any exception to this must be approved by the Graduate Director.**

Course Descriptions:

Course: STT 863 Statistical Methods I
Prerequisite: STT 442 or STT 862, MTH 415
Offered: Fall, every year

Description: Introduction to the general theory of linear models; Application of regression models; Interval estimation, prediction and hypothesis testing; Contrasts; model diagnostics; model selection; LASSO type algorithm; Introduction to Linear mixed effect models.

Detail Description:
Recommended Text Book: *Applied Linear Regression Models by Kutner* (4th Edition), Nachtsheim and Neter, McGraw-Hill.

The emphasis of this course is linear regression models, which are widely used in business administration, economics, engineering, and the social, health, and biological sciences. Successful applications of these models require a sound understanding of both the underlying theory and the practical problems that are encountered in using the models in real-life situations. While this is basically an applied course, it seeks to blend theory and applications effectively, avoiding the extremes of presenting theory in isolation and of giving elements of applications without the needed understanding of the theoretical foundations. The topic details are as follows.

Linear regression with One predictor variable; Functional relation between two variables; Statistical relation between two variables; Basic concepts of regression models; Construction of regression models; uses of regression models; regression and causality; Simple linear regression model with distribution of error terms unspecified; formal statement of model; important feature of model; meaning of regression parameters;

Inference: Sampling distribution of regression coefficient estimators, estimated means and predictors; Confidence intervals and testing of regression coefficients, means and new observations; Illustrative examples and additional results. Analysis of Variance approach to Regression analysis; Partitioning the SS and d.f; use of ANOVA table for testing; Full model; Reduced model; Coefficient of Determination.

Diagnostics: Diagnostics for predictor variables; residuals; semi and studentized residuals; Graphical and analytical way of testing nonlinearity; nonconstant error variance; non-normality; Brown-Forsythe Test; Breusch Test;

Remedial Measures: nonlinearity of regression; nonconstancy of error variance; nonindependence of error; omission of important variables; Transformations; Transformation for non-linearity; transformation for non-constancy and non-normality; Box-Cox Transformation; Exploration of Shape of Regression Function; Lowess method; Use of smoothed curves to confirm fitted regression function.

Simultaneous inference of regression coefficients; Bonferroni and Working-Hotelling Procedures; simultaneous prediction intervals; introduction of measurement error models and inverse prediction.

Mutiple Linear Regression; Need of multiple regression models; parameter interpretations; estimation and testing of different parameters; partial F test; partial coefficient of determination; multicolinearity and other diagnostics. Ploynomial regression, interaction regression models; qualitative predictors

Model selection and validation; criteria for model selection; automatic search procedure; LASSO.

Linear mixed models: Gaussian mixed models through some Illustrative examples.

Course: STT 864 Statistical Methods II
Prerequisite: STT 863
Offered: Spring of every year

Description: Generalized linear models(GLMs); Deviance and residual analysis in GLMs; Analysis of two-way and three-way contingency tables; Logistic regression; Log-linear models; Multicategorical response models; Poisson regression; Survival analysis; Introduction to generalized estimating equations; Introduction to longitudinal data; Bayesian analysis using WinBUGS

Detailed description:
Recommended Text: *An Introduction to Generalized Linear Models (3rd edition)* by Dobson and Barnett, CRC Press

Introduction to GLMs, exponential family distribution, the random and systematic components of a GLM, the (canonical) link function, Properties of distributions in the exponential family; maximum likelihood estimation; inference; Binary variables and logistic regression; dose response models; goodness-of-fit statistics; residuals; diagnostics. Nominal and ordinal logistic regression; multinomial distribution; nominal logistic regression; ordinal logistic regression; Poisson regression and log-linear models. Introduction to survival analysis; survival function and hazard models; Cox regression models; statistical estimation and inference. Introduction to clustered and longitudinal data, (basic models only). Bayesian analysis; priors; distributions and hierarchies in Bayesian analysis; use of WinBUGS software; Markov chains, Bayesian inference, convergence diagnostics (no theory). Examples of Bayesian Analysis in logistic regression, survival models, random effects models and longitudinal data analysis.

STT 867: Linear Model Methodology
Prerequisite: STT 862 or equivalent
Offered: Fall, every year

Text: Linear Models by Searle (Wiley Classics Library)
   Linear Model Methodology by Khuri (CRC Press)
   An Introduction to Categorical Data Analysis by Alan Agresti

Course Description: Properties of multivariate normal distribution, Cochran's theorem, Simple and multiple linear regression models, Gauss-Markov Theorem, Theoretical properties, BLUP, one-way and two-way ANOVA models, SS's, diagnostics and model selection , contingency tables and multinomial models, generalized linear models, logistics regression.

Course content:
1. Distributions and Quadratic forms (2 weeks)
   Normal (univariate and multivariate), t, F, Chisquare (central and non central), and Wishart distributions; Quadratic forms in normal variables, Cochran's theorem.
2. Regression, Full rank models (3/4 weeks)
   Simple linear regression model, estimation, examples, multiple linear regression; ols, gls, ML, BLUE; properties of the estimators; Analysis of Variance; general linear hypothesis, examples.
3. Regression on Dummy Variables and polynomial regression (1 week)
4. One-way and two-way ANOVA models with examples (1.5 weeks)
5. Diagnostics and model selection (1.5 week)
6. Introduction of 2X 2 contingency tables, multinomial models and statistical inference (2 weeks)
7. Generalized linear models, likelihood formulation, Weighted iterative algorithm, inference; analysis of deviance (1.5 week)
8. Logistic regression, model interpretation, statistical properties (1.5)
9. Large and small sample inference for generalized linear models.

STT 868: Mixed models: Theory, Methods and Applications
Prerequisite: STT 862 & 867 or equivalent
Offered: Spring, every year

Course Description: Introduction and motivation; maximum likelihood estimation and other estimation methods for linear mixed models; statistical properties of the LME models; prediction under LME models; generalized linear mixed models; quasi-likelihood estimation, generalized estimating equations for GLMM; nonlinear mixed models; diagnostics and influence analysis; Bayesian development in mixed linear models (no detail); application in longitudinal data analysis, growth curve analysis, epidemiology, genetic data and health data.

Text: *Generalized, Linear and Mixed Models (2nd edition)* by McCulloch and Searle and Neuhaus, Wiley, 2008.

References:
Mixed Models: Theory and Applications by E. Demidenko
Linear and Nonlinear Models for the Analysis of Repeated Measurements by E.F. Vonesh and V.M. Chinchilli.
Nonlinear Models for Repeated Measurement Data by M. Davidian and D.M. Giltinan.
Linear Mixed Models for Longitudinal Data by G. Verbeke and G. Molenberghs
Analysis of Longitudinal Data by P. Diggle, P. Heagerty, K.Y. Liang and S.L. Zeger.

Linear Mixed Models for discrete Longitudinal Data by G. Verbeke and G. Molenberghs

STT872: Statistical Inference I
Prerequisite: STT 862 & 881
Offered: Spring, every year

Course Description: Statistical distributions, Decision theoretic formulations of estimation and testing of hypothesis, Sufficiency, Rao-Blackwellzation, admissibility, Bayes and minimax estimation, maximum likelihood estimation, inference based on order statistics, Neyman-Pearson lemma and applications, multiple testing, linear models and Gauss-Markov Theorem.

Course content: Review of some distributions (.67 week): Normal, Gamma, bets, t and F distributions. Decision theoretic formulations of estimation and testing of hypothesis (1wk), Sufficiency, Rao-Blackwellzation (1.5wks), admissibility, Bayes and minimax estimation (2.5wks), maximum likelihood estimation (classical consistency and asymptotic normality proof)(1.33wks), inference based on order statistics (1.5wks), Neyman-Pearson lemma and applications (2wks), multiple testing (Scheffe's method, Tukey's, FWER)(2wks), linear models and Gauss-Markov Thm (2wks).

Texts: This course can be drawn from the books of Lehmann on estimation and testing; C.R. Rao on Inference for Linear Models;

STT 873: Statistical Learning and Data Mining
Prerequisite: STT 868 & 872 or equivalent
Offered: Fall, odd years

Course Description: Statistical methods focusing on machine learning and data mining, modern regression and classification techniques, Support Vector Machines, Boosting, kernel methods and ensemble methods, clustering, dimension reduction, manifold learning, and selected topics.

Text: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* by T. Hastie, R. Tibshirani, and J. Friedman.
Reference: *Introduction to Statistical Learning Theory* by O. Bousquet, S. Boucheron, and G. Lugosi.

Course content:
1. Linear regression and related methods (3.5 weeks): Shrinkage, ridge regression, principal components regression, partial least square regression. Model selection, least angle regression and the lasso, stepwise methods. Cross-validation and bootstrap.
2. Classification (2 weeks): Linear discriminant analysis, logistic regression, and support vector machines (SVMs). Surrogate loss functions and fisher consistency.
3. Nonparametric regression and classification (5.5 weeks): Basis expansions, splines and regularization. Kernel methods. Kernel smoothing. Generalized additive models. Coordinate gradient descent algorithm. Classification and regression trees. Ensemble methods. Multivariate adaptive regression splines. Bagging. Random forest. Boosting, additive trees and greedy algorithms. Neural networks. Prototype & near-neighbor methods.
4. Unsupervised learning (2.5 weeks): Clustering. Gaussian mixtures and the EM algorithm. Linear and nonlinear dimension reduction. Manifold learning.
5. Selected topics (1.5 week): High-dimensional learning, large scale multiple testing

STT 874: Introduction to Bayesian Analysis
Prerequisite: STT 868 & 872 or equivalent courses

Offered: Fall, even years

Course Description: A Bayesian method provides statistical inference by combining priori information and the observations. Its popularity has been increased due to the development of Markov chain Monte Carlo (MCMC) methods and the advances in computing power. The Bayesian method are widely used in various applied areas such as biological science, public health, epidemiology, education, social sciences, agriculture, engineering, etc. This course will introduce the basic theory of Bayesian frameworks including empirical Bayes (EB), hierarchical Bayes(HB) and nonparametric Bayes(NP), computational methods for Bayesian inference such as a Gibbs sampler, a Metropolis-Hastings method, etc. and applications with real data examples from various fields.

Suggested Text: *Bayesian Data Analysis* by Gelman, Carlin, Stern and Rubin

Details materials to be covered:
Basic Theory in Bayesian methods: 6 weeks
-Bayes Theorem, Prior, Posterior, Predictive Distributions
-Bayesian Inference including point estimation, interval estimation, Decision theory
-Empirical Bayes, Hierarchical Bayes, Bayesian asymptotic methods
-Nonparametric Bayes including Dirichlet process, Dirichlet Mixture, Stick-breaking Processes, Polya Tree process, etc.
Computation methods: 8 weeks
-Basic numerical methods, sampling methods
-Markov Chain theory
-MCMC algorithms including Gibbs sampler, M-H algorithm and their variants, RJMCMC for both parametric, nonparametric Bayesian methods
-Diagnostic criterion using MCMC samples, model selection and validation
Applications with various statistical models : 2 weeks
-Density estimation, parametric/nonparametric regression problems
-Linear mixed models, generalized linear mixed models
-spatial regression, prediction and area-level modeling
-clustering, classification, functional data analysis

STT881: Theory of Probability I
Prerequisite: STT 861 & MTH 421
Offered: Fall, every year

STT881 is on measure and integration theory and limit theorems for independent random variables. It covers the following topics.

Course Description: Measures and their extensions, integration. Lp spaces and Inequalities. Lebesgue decomposition, the Radon-Nikodym theorem. Product measures, Fubini's theorem, Kolmogorov consistency theorem. Independence, Kolmogorov's zero-one law, the Borel-Cantelli lemma. Law of large numbers. Central limit theorems, characteristic functions, the Lindeberg-Feller theorem, asymptotic normality of sample median. Poisson convergence. Conditional expectations.

STT882: Theory of Probability II
Prerequisite: STT 881
Offered: Spring, every year

STT882 is on theory of stochastic processes. It covers the following topics.

Course Description: Random walks, transience and recurrence. Martingales, martingale convergence theorem, Doob's inequality, optional stopping theorem. Stationary processes and Ergodic theorem. Brownian motion, Kolmogorov's continuity theorem, strong Markov property, the reflection principle, martingales related to Brownian motion. Weak convergence in C([0,1]) and D([0, 1]), Donsker's invariance principle, empirical processes.

These two courses will use the following two textbooks:
R. Durrett, *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics, 2010.

K. B. Athreya and S. N. Lahiri, *Measure Theory and Probability Theory*. Springer, 2006.

STT 951: Statistical Inference II
Prerequisites: STT 872, 882
Offered: Spring, odd years

Course description: Decision theoretic estimation: Minimaxity, admissibility, shrinkage estimators, James-Stein estimators. Advance estimation theory, invariant tests, multiple testing, FWER, FDR, and related methods. Invariance and testing of hypotheses: Maximal invariant, most powerful invariant test, permutation and rank tests, unbiasedness and invariance, max-min test, Hunt Stein theorem.

Contents of the topics here could be drawn from chapters 5 and 6 of the book by Lehmann on *Theory of Point estimation, Second Edition*, 1998; chapters 6, 8, and 9 of the book by Lehmann and Romano on *Testing Statistical Hypotheses, Third Edition*, 2005; the monograph by Efron on *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction,* 2010, Cambridge University Press, and from various papers.

STT 953: Asymptotics theory
Prerequisites: STT872 & 882
Offered: Spring, even years

Description: Locally asymptotic normal models, empirical likelihood, U-statistics, asymptotically efficient and adaptive procedures.

STT 961: Weak Convergence and Asymptotic Theory
Prerequisite: STT 872, 882
Offered: Fall, odd years

Course description: Maximal inequalities, covering numbers, symmetrization technique, Glivenko-Cantelli Theorems, Donsker Theorems and some results for Gaussian processes, Vapnik-Chervonenkis classes of sets and functions, applications to M-estimators, bootstrap, delta-method

Text: *Weak convergence and Empirical Processes with applications to statistics* by A. van der Vaart and J. Wellner.

Course content: Maximal inequalities (1.5 weeks), covering numbers (1.5 weeks), symmetrization technique (1.5 weeks), Glivenko-Cantelli Theorems (1.5 weeks), Donsker Theorems and some results for Gaussian processes (1.5 weeks), Vapnik-Chervonenkis classes of sets and functions (2 weeks), M-estimators (1.5 weeks), bootstrap (2 weeks), delta-method (2 weeks)

STT 962: Fractional Processes and Power Laws
Prerequisite: STT 872, 882
Offered: Spring, even years

This course is an introduction to self-similar processes and random fields with long-range dependence and/or heavy-tailed distributions. It covers selected topics from the following list.

Course Description: Self-similar processes. Fractional Brownian motion, fractional stable motions. Fractional calculus, Laplace and Fourier transforms, semigroups and generators. Continuous time random walks. Connections between long range dependence, heavy tails, and fractional calculus. Inference for processes with long range dependence and heavy tails, including fractional ARIMA models, ARCH/GARCH models, and random difference equations.

This course could be based on the following books:

Samorodnitsky & Taqqu. *Stable non-Gaussian random processes. Stochastic models with infinite variance*, 1994, Chapmann & Hall.

Meerschaert & Sikorskii. *Stochastic Models for fractional calculus*, 2012, De Gruyter,

Cohen & Istas. *Fractional Fields and Applications.* 2012, Springer-Verlag.