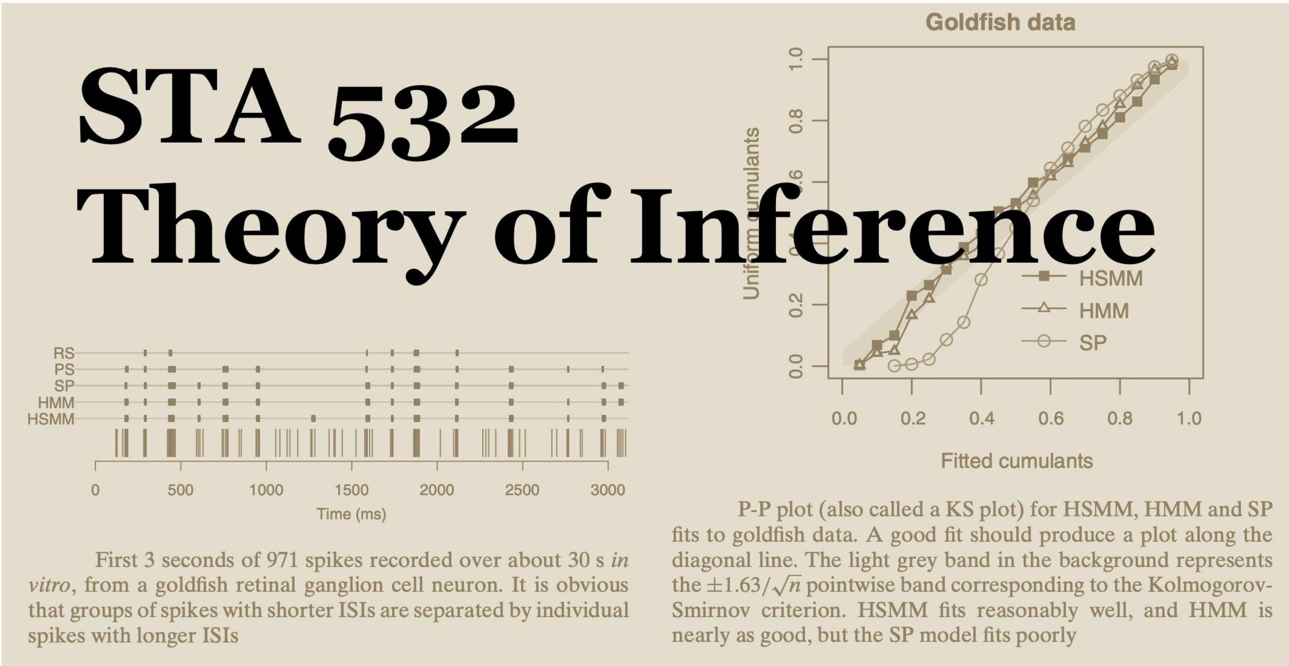
Send To Printer | Close Window

Syllabus



First 3 seconds of 971 spikes recorded over about 30 s *in vitro*, from a goldfish retinal ganglion cell neuron. It is obvious that groups of spikes with shorter ISIs are separated by individual spikes with longer ISIs

P-P plot (also called a KS plot) for HSMM, HMM and SP fits to goldfish data. A good fit should produce a plot along the diagonal line. The light grey band in the background represents the $\pm 1.63/\sqrt{n}$ pointwise band corresponding to the Kolmogorov-Smirnov criterion. HSMM fits reasonably well, and HMM is nearly as good, but the SP model fits poorly

What this course is about

This is a master's level course on the theory of statistical inference. To some statistics is a collection of numerical recipes for data analysis. To others statistics is a branch of applied mathematics for taking decisions under uncertainty. These views lead to the ***whats*** and ***hows*** of the practice of statistics. What elevates statistics to an academic discipline of its own is a third perspective which gets to the bottom of the ***whys***. Why do we do what we do? Why does it matter how it is done? Why do Bayesians obsess over posteriors? Why do frequentists fret over priors? Why do we need to learn both paradigms? In this course we will primarily focus on the whys.

Text

These lecture notes will serve as the main text for the course. Supplement them with Casella and Berger (Statistical Inference) and Wasserman (All of Statistics)

We will conclude with this article on The Interplay of Bayesian and Frequentist Analysis by Susie Bayarri and Jim Berger (Statistical Science, 2004)

Tentative Schedule

Jan 21, 26, 28, Feb 2, 4, 9, 11: Probability Review   HW1, HW2

Feb 16, 18, 23, 25: Statistical Inference, Classical Inference HW3, HW4
Mar 02, 04: Review
Mar 9: No Classes
Mar 11: Midterm Exam
Mar 16, 18, 23, 25, 30, Apr 01: Classical Methods HW5, HW6
Apr 06, 08, 13, 15: Optimal Learning HW7
~~Apr 13, 15: High Dimensional Inference~~
~~Apr 20, 22: Bayes Frequency Milieu~~
Friday April 30 9 AM - Noon: Final Exam

Weekly homework assignments starting from first full week. Submissions due Sunday night. Solution roadmaps will be posted after resubmission is done.

Lecture Meetings

This course meets remotely via Zoom. Lectures are live and students are encouraged to attend and have their videos switched on for an interactive experience. However, all lectures will be recorded and made available on Sakai for later viewing. All assessment (homework and exams) will be administered via Sakai.

Lectures: TTh 8:30-9:45 AM US Eastern Time

Office Hours: TBD

Instructional Team

Instructor
Surya T Tokdar, Associate Professor of Statistical Science
219A Old Chemistry Building
Email: surya.tokdar@duke.edu
Web: www.stat.duke.edu/~st118/
Phone: 9196842152

Teaching Assistants
Christine Shen, Statistical Science MS Student (2nd year)
Email: yueming.shen@duke.edu

Shounak Chattopadhyay, Statistical Science PhD Student
Email: shounak.chattopadhyay@duke.edu

Exam Dates

Midterm Exam: Thursday 3/11
Final Exam: Friday 4/30

Course Description

This is a master's level course on the theory of statistical inference. To some statistics is a collection of numerical recipes for data analysis. To others statistics is a branch of applied mathematics for taking decisions under uncertainty. These views lead to the ***whats*** and ***hows*** of the practice of statistics. What elevates statistics to an academic discipline of its own is a third perspective which gets to the bottom of the ***whys***. Why do we do what we do? Why does it matter how it is done? Why do Bayesians obsess over posteriors? Why do frequentists fret over priors? Why do we need to learn both paradigms? In this course we will primarily focus on the whys.

Text

These lecture notes will serve as the main text for the course. Supplement them with Casella and Berger (Statistical Inference) and Wasserman (All of Statistics).

Prerequisite

STA 601/602 and STA 521. We need multivariable calculus, matrix algebra, basic real analysis, calculus based probability, explosure to regression analysis or equivalent applied statistics course, fluency with Bayesian modeling and computing.

Assessment

The course will have 12 to 13 weekly homework assignments (50%), a midterm exam (30%) and a final exam (20%). Graded work on homework assignments and midterm exam could be revised and resubmitted to recover missed points at 50% rate.

Policies

We observe strict adherence to Duke Community Standard. No permission is needed to miss a lecture once in a while, but prior communication on a planned miss is appreciated. Prior notification is required for longer stretches of missed lectures. Missing homework should be notified and make-up should be requested prior to submission deadline. No make up allowed to either midterm of final exam. Missed midterm will be subsumed by the final exam.

Students are allowed to discuss homework assignments. But everyone must submit their own original work. No discussion or collaboration is allowed for either midterm of final exam. Any academic misconduct will be reported to appropriate university office.

Procedure for Testing Accommodations

This class will use the Testing Center to provide testing accommodations to undergraduates registered with and approved by the Student Disability Access Office (SDAO). The center operates by appointment only and appointments must be made at least 7 consecutive days in advance, but please schedule your appointments as far in advance as possible. You will not be able to make an appointment until you have submitted a Semester Request with the SDAO and it has been approved. So, if you have not done so already, promptly submit a Semester Request to the SDAO in order to make your appointment in time. For instructions on how to register with SDAO, visit their website at https://access.duke.edu/requests. For instructions on how to make an appointment at the Testing Center, visit their website at https://testingcenter.duke.edu.

Statistical Inference

Surya Tokdar

Spring 2021

Contents

Preface

This is a master's level course on the theory of statistical inference. To some statistics is a collection of numerical recipes for data analysis. To others statistics is a branch of applied mathematics for taking decisions under uncertainty. These views lead to the *whats* and *hows* of the practice of statistics. What elevates statistics to an academic discipline of its own is a third perspective which gets to the bottom of the *whys*. Why do we do what we do? Why does it matter how it is done? Why do Bayesians obsess over posteriors? Why do frequentists fret over priors? Why do we need to learn both paradigms? In this course we will primarily focus on the whys.

Of course a discussion of whys would be useless without knowing some of the whats and hows. I will assume the reader is reasonably knowledgable of the Bayesian approach to statistical data analysis. You know how to fit and summarize basic conjugate and non-conjugate Bayesian models, and have had some exposure to slightly more complex hierarchical models. Additionally, I will assume you have been introduced to regression models and analysis (either Bayesian or least-squares based).

Implicit in these assumptions is a more basic one: you have studied probability at a level requiring multivariable calculus. You are comfortable manipulating probability mass and density functions of one or several variables, including the somewhat subtle concept of conditioning on a continuous random variable. I will also assume you are comfortable with basic matrix algebra (basic manipulation, ranks, determinants, traces, inversion, projection etc.). Chapter 1 gives a comprehensive review of the probability background we need. We will spend two to three weeks to cover these in some detail.

On the other hand, I will *not* assume you have seen a whole lot of classical or non-Bayesian methods. Chapters 3 and 4 are dedicated to the whats and hows of classical statistics. But we will continue to learn more about these in the remaining chapters. However, this is not the course where you learn all the alphabet tests (t, F, χ^2 and so on) or ANOVA or how to fit a quantile regression model. We will learn broad principles and see some examples for further illumination. You should take an applied statistics course if you were more keen on the whats and hows.

As we progress, we will dive deeper into complex scenarios where the fundamental theory questions will take a fascinating shape (Chapters 6 in particular), bringing forth a new set of whats and hows. We will conclude our discussion with an ultimate how question: How does statistical practice benefit from accepting a milieu of both Bayesian and frequentist thinking?

These lecture notes, written in a book style, are purposely more comprehensive than the typical class handout. It is important for a course on theory to have some breadth and depth so that you have more things to read and think about than what I will be able to go over in the lectures. Not all of the additional reading will be useful. But my hope is that in every chapter, you will find an interesting question or thought that you can then pursue on your own to expand your knowledge and critical thinking.

Typical text books on the theory of inference, presenting essentially the frequentist paradigm, do not fit our bill. Graduate programs at Duke Statistical Science revolve around applied Bayesian modeling and computing. It is my job to convince you through this course that while you might remain Bayesian in your practice, you need to understand both Bayesian and frequency based foundations of statistical inference and perhaps, even better, to consider the possibility that their synthesis is crucial for statistical science to be useful in science and society.

Each of Chapters 1 through 7 concludes with a comprehensive set of exercise problems that are mostly longer and more verbose than your typical text book examples. I believe much of the learning will take place by working out these exercise problems and thinking deeply about related issues. This course will require a high level of active engagement from you to make most out what these notes have to offer.

Finally, these set of notes are a work in progress. I will tremendously appreciate feedback, proofreads, error corrections and suggested edits and expansions to improve them. If this collection ever comes out as a book, every contributor will be duly acknowledged and thanked from the bottom of my heart!