Term: Spring 2022

Course Instructor: Tracy Ke <zke@fas.harvard.edu>

Teaching Fellows: Qizhao (Jane) Chen <qizhaochen@g.harvard.edu>

Meeting Time: 1:30-2:45 PM, Tuesday/Thursday

Office Hours (Zoom): 4:00-5:00PM Monday (Tracy); 7:00-8:00PM Wednesday (Jane)

Office Hours during the shopping period (Eastern time):

- Jan 12 (Wednesday) 2:00-3:00 PM
- Jan 18 (Tuesday) 2:00-3:00 PM

The Zoom links can be found from STAT 131/Zoom.

Grading: Homework (30%), midterm (25%), written final (25%), data-analysis final (10%), attendance & quizzes (10%)

Attendance & quizzes: There will be a short quiz (usually 10 minutes) in the beginning of each Tuesday's lecture (starting from Feb 1). The quiz serves as both a check-in of attendance and a test of basic things taught in class. Sample quiz will be posted during the shopping period.

Midterm: Close-book in-class exam (1 cheatsheet allowed). No R coding.

Written final: In the exam period. No R coding.

Data-analysis final: Open-book, open-notes, in the last lecture. There are only data analysis questions (these questions are similar to those appearing in homeworks.) At the end of the exam, each student should submit R code and output. For questions of this data-analysis final, please come to the office hours during shopping period.

Homework: A mix of mathematical statistics (how to formalize real problems, come up with an estimator, prove consistency, etc.) and implementation of the methods introduced in lecture (using R or the language of your choice; solutions will be provided in R). There will be will be approximately 8 or 9 assignments over the semester.

Late Homework Policy: Late assignments will automatically receive a grade of zero. NOTE: There will be NO opt-out of 1 homework this year.

Course collaboration policy:

For quizzes: No collaboration is allowed.

For homework: You are allowed and in fact encouraged to discuss homework with other students, but all of your final answers should be in your own words. Copying or simply paraphrasing other studentsâ€™ homework is not allowed. Please list all the the other students that you have discussed your assignment with at the front of your submitted homework (or as a comment to your online submission). In particular, you can ask questions and seek help from each other, but you are not allowed to copy others' code/plots/proofs.

Scenario 1: Student A asks Student B how to use an R function. Later, A implements it by himself/herself and generates the results/plots needed for the homework. [This is Okay.]

Scenario 2: Student A takes the homework of Student B, copies the lines of code and/or the results/plots. [This is not allowed.]

Scenario 3: Students A&B discuss about how to prove something in the homework. They figure out the answers together. Later, each of them writes down the solutions independently. [This is Okay.]

Scenario 4: Students A&B discuss about how to prove something in the homework. They figure out the answers. Student B writes down all steps. Student A copies his/her answers. [This is not allowed.]

For general course materials: You are encouraged to discuss lecture materials with other students or form study groups.

For exams: All exams (including the midterm and finals) should be your independent work - no

collaboration is allowed.

Course Description: This course is an introduction to time series with an emphasis on models, methods, and algorithms. Time series techniques are used to understand data which may be observed over time and in which time plays an important role.

We will begin with simple models such as autoregressive, moving average, and ARIMA models, which involve linearity and Gaussianity assumptions. With these models in mind, we discuss statistical questions such as parameter inference, model choice, and forecasting in the presence of uncertainty.

In the second part of the course we will cover more sophisticated time series models such as hidden Markov and state space models, which are used extensively in science and industry. In turn, these suggest interesting computational questions, and we will cover classic algorithms such as the Kalman filter and particle filter.

Course Requirements: No previous knowledge of time series methods is required, but we assume knowledge of statistical inference and probability at the level of STAT 110, 111, and 139, or the equivalent in other departments. For example, we will frequently use concepts such as *likelihood*, *estimators*, *asymptotics*, *variance and covariance*, *method of moments*, *conditional distributions*, etc.

Some reminders on calculus and probability, written by Prof. Neil Shephard who previously taught this course, are available in Files/Reading Material. The topics mentioned above are also covered in the first seven chapters of *All of Statistics* by Larry Wasserman. If you plan to take the course, these should not look too scary to you.

Text: We will upload slides and lecture notes regularly, so a textbook is not strictly necessary.

For reference, the closest book for the first part of the course will be *Time Series Analysis and Its Applications* by Shumway and Stoffer (*PDF can be found under STAT131/Files*). For the second part the course will follow the themes of *Nonlinear Time Series* by Douc, Moulines, and Stoffer, without being as heavy on the math side.

Other standard books on time series include *Introduction to Time Series and Forecasting* by Brockwell and Davis (1991), *Financial Time Series* by Tsay (2010), and *Time Series Analysis with Applications in R*, by J.D. Cryer and Kung-Sik Chan (2008).

Goals

- Foundations: Specificities of dependent data (i.e. what makes time series special?); the logic underlying statistical inference in this setting; interpretation of parameters, model criticism; asymptotic properties in the limit of the number of data points.
- Skills: Calibration and prediction of time series; implementation in R on various examples; uncertainty quantification; algorithms and computational methods for hidden Markov models and state space models.

Weekly agenda is here: Agenda-Spring-2022.pdf