Statistical and Computational Methods for Learning through Graphical Models

Instructor: Zhenke Wu PhD, Assistant Professor of Biostatistics
Email: zhenkewu@umich.edu
Time: Tuesday and Thursday 12:30-2pm
(15 weeks; September 6th to December 13th, 2016)
Location: 4332 SPH II
Office Hours: Tuesdays 2-3pm and by appointment
Course Website: http://zhenkewu.com/teaching/graphical_model

"Graphical statistical models play an increasingly important role in epidemiology and in the translation of epi research to the health of the individual. The information and biotechnology revolutions of the past 2 decades are generating novel, complex health measurements, as well as the analytic power to use them to better infer an individual's health status and trajectory and the likely benefits and harms of particular health decisions, such as choice among competing treatments. We formalize a particular representation of the problem of individualized health by way of a basic graphical model that demonstrates how population studies can inform and be informed by individual health decisions and outcomes."

– Ogburn and Zeger (2016), American Journal of Epidemiology

"The formalism of probabilistic graphical models provides a unifying framework for capturing complex dependencies among random variables, and building large-scale multivariate statistical models. Graphical models have become a focus of research in many statistical, computational and mathematical fields, including bioinformatics, communication theory, statistical physics, combinatorial optimization, signal and image processing, information retrieval and statistical machine learning."

– Wainwright and Jordan (2008), Foundations and Trends in Machine Learning

Dear Students in Advanced Topics in Biostatistics,

Thanks for choosing to learn the language of graphical models and their applications this semester. This note explains the course objectives, organization, and method of evaluation along with other information to get you ready for this course. We are excited about graphical models and about supporting your learning about them.

Research on graphical models is experiencing unprecedented growth, largely stimulated by the analytic tasks required for "big data" defined (by IBM) by their large "volume, variety, velocity and veracity". The course is structured to offer students the foundations, tools, and latest research results to navigate or even expand this literature.

Course Objectives:

- To familiarize students with the concepts, applications and computational techniques of graphical models.
- To engage students in building, estimating and interpreting expert systems for problems either suggested by the instructor or identified by the students.
- To showcase the current frontier of graphical model research in biomedical problems and to prepare advanced PhD or Masters students for their next research projects.

Session Organization – The course is organized into four modules: (1) graphical representation of probability distributions; (2) inference and fast computation given observed data; (3) graphical models for causal inference and, (4) relevant case studies that serve to review the ideas and methods and introduce students to active research topics.

Module 1 (3 weeks): Representation

1. Graph structure and terminologies; Why study graphical models?
2. Directed graphical models
3. Undirected graphs models
4. Other variants of graphical models

Module 2 (4 weeks): Inference and Computation for Graphical Models

1. Exact algorithms
2. Approximate algorithms
3. Scalable Bayesian algorithms
4. Structure learning
5. Software packages

Module 3 (3 weeks): Graphical Models for Causality

1. Causal graphical models: concepts and inference
2. Structure learning of causal graphs
3. Causal inference for network data (randomization; peer-encouragement design, etc.)

Module 4 (4 weeks): Case Studies

1. Individualized health problems (partially-latent class models, dynamic Bayesian networks, etc.)
2. Large-scale networks (latent state space models)
3. Deep learning examples

4. Graphical models for neuroimaging data (Guest lectures, TBD)

Optional topics:

5. (Optional) Regression tree models for heterogeneous treatment effects, survival data
6. (Optional) Parametrization and Inference of discrete graphical models

Course Materials:

The lectures will use materials from books and foundational research papers relevant to specific topics. New developments will be introduced via required and optional paper readings. The Chapter numbers and papers will be provided at the end of each lecture.

Books

1. Koller, Daphne, and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
2. Wainwright, Martin J., and Jordan, Michael I.. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning* 1, no. 1-2 (2008): 1-305. (https://people.eecs.berkeley.edu/~wainwrig/Papers/WaiJor08_FTML.pdf)
3. Cowell, Robert G., ed. *Probabilistic networks and expert systems: Exact computational methods for Bayesian networks*. Springer Science & Business Media, 2006.
4. Pearl, Judea. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009.

Required papers – The papers that introduced key concepts in graphical modeling, designed new inferential algorithms or inventively applied graphical models to important problems.

Optional papers – Two kinds of papers: 1) Elaborations and Illustrations of key concepts. 2) Advanced papers that illustrate cutting-edge research on relatively specific topics. Students can use them for motivating new research.

Students will be evaluated with the following weighting

Homework Assignment - 30%. (The total homework grade equals the sum of 3 highest scores out of four, each corresponding to one learning module and graded in the scale of 0-10.) The homework will be assigned one week prior to the end of each module. Assignments will be due 1 week after the module completion.
Active participation - 10%. Peer-review. Help oneself learn and teach one's classmates and instructor by asking questions and discussing solutions.

Term Project – 60% (Poster presentation on ***December 13th, 2016***) Based on the trimmed mean of the scores obtained from external judges and the instructor. A ***separate***, but ***optional*** report will be due at 11:59pm ***December 20th, 2016.*** Students with ONLY poster presentation will be graded solely on poster scores; those with ADDITIONAL written report will be graded based on the LARGER of the two: the poster and the written report scores.

Get ready for class:

Academic Integrity - The faculty and staff of the School of Public Health believe that the conduct of a student registered or taking courses in the School should be consistent with that of a professional person. Courtesy, honesty, and respect should be shown by students toward faculty members, guest lecturers, administrative support staff, community partners, and fellow students. Similarly, students should expect faculty to treat them fairly, showing respect for their ideas and opinions and striving to help them achieve maximum benefits from their experience in the School. Student academic misconduct refers to behavior that may include plagiarism, cheating, fabrication, falsification of records or official documents, intentional misuse of equipment or materials (including library materials), and aiding and abetting the perpetration of such acts. Please visit http://sph.umich.edu/student-resources/mph-mhsa.html for the full Policy on Student Academic Conduct Standards and Procedures.

Student Accommodations - Students should speak with their instructors before or during the first week of classes regarding any special needs. Students can also visit the Office of Academic Affairs for assistance in coordinating communications around accommodations. Students seeking academic accommodations should register with Services for Students with Disabilities (SSD). SSD arranges reasonable and appropriate academic accommodations for students with disabilities. Please visit https://ssd.umich.edu/topic/our-services for more information on student accommodations. Students who expect to miss classes, examinations, or other assignments as a consequence of their religious observance shall be provided with a reasonable alternative opportunity to complete such academic responsibilities. It is the obligation of students to provide faculty with reasonable notice of the dates of religious holidays on which they will be absent. Please visit http://www.provost.umich.edu/calendar/religious_holidays.html#conflicts for the complete University policy.

Diversity - SPH is committed to creating classroom environments that are supportive of diversity, equity and inclusion.

Again, thanks for participating in Advanced Topics in Biostatistics. We look forward to working and learning together this term.

Zhenke Wu PhD, Assistant Professor of Biostatistics and Instructor