Iowa State University – BCB 568 Syllabus – Spring 2015

Instructor: Karin S. Dorman
Office: Snedecor 2411; Science II 534
Phone: (515)294-1457; (515)294-6078
Email: kdorman@iastate.edu
Office Hrs: Contact me.
TA: Xin Yin
xinyin@iastate.edu
Snedecor 3219

Lecture: TR 9:30 - 10:50
Location: MBB 1424
Webpage: http://thirteen-01.stat.iastate.edu/wiki/bcb568/
(Ask instructor for username/password.)

Course Description


Prerequisites. STAT 430 or equivalent introduction to probability and statistical notation and thinking.

Course Goals

Learning Outcomes.

• Comprehend and critique bioinformatics articles that include statistical models and arguments.
• Formulate models to answer basic bioinformatics questions.
• Recognize and account for uncertainty and noise in biological sequence data.
• Recognize and account for dependence in biological sequence data.
• Use a model and standard statistical tools to perform inference: estimate model unknowns, answer scientific questions about the model, assess confidence in results, etc.

Learning Objectives.

• Given an manuscript with a probability or statistical model, read, comprehend, correct, and explain the model to colleagues.
• Given sequence data containing multiple instances with variation of an unknown short motif, formulate a model, write a likelihood equation, and use it to find the motif and its occurrences.
• Given next generation sequencing (NGS) data aligned to a reference genome, identify the sequencing errors, the sequence variants, and assess confidence in each conclusion.
• Given NGS data aligned to a reference genome, test thousands of genes for differential expression and produce a list of genes such that at least 90% are really differentially expressed.

Course Information

Textbook: There is no required textbook for this class. You will be given all materials needed for this class. Nevertheless, you may find the following books useful references:
Primary literature: We will read one or a few articles from the bioinformatics and statistical literature nearly every week. Some will be current publications, others will be seminal work, foundational to modern statistical bioinformatics.

Class Participation: You will spend the rest of your professional life self-learning from books and MOOCs, but the empirical evidence shows that in-class experiences encourage better and broader retention. You will be required to attend and participate in class. Some days there will be reading assignments. You are expected to complete the reading before class in order to participate in discussion or in class assignments. Your performance in class will be judged at the end of the semester by your instructor and peers. You will be rewarded for asking helpful or insightful questions, for helping your colleagues solve in-class problems, and for finding mistakes or clarifying content delivered as part of the class.

Absence: Please let the instructor know if you must be absent. The instructor will work with you if illness or other serious issues interrupt your work or take you away during critical times, but you must keep her informed of your status.

Grading: Letter grades including plus/minus will be given based on performance on homeworks (30%), quizzes/exams (30%) a project (30%), and in class participation (10%). The specific grading scale is not determined until after all grades have been calculated. You will get periodic feedback about your projected grade based on completed work, but keep in mind a large fraction of the points are earned or assessed near the end of the class (last homework, project, final exam, and participation).

Homework: Individual practice (homework) is an essential part of learning. Homework assignments will be due by midnight on the due date. Solutions will be posted on the course website.

Midterm: The midterm exam will be given during regular class time. If you have a conflict, you must let me know before the exam. Failure to do so will result in a 0 for the exam.

Final Exam: The final exam is cumulative with special emphasis on the material covered in the second part of the course. The time for the exam is posted above. Do not make plans for the end of the semester that prevent you from attending the final exam.

Project: You will choose a paper (or few papers) with a statistical component around which you will design a lesson in statistical bioinformatics, including background reading, a lesson plan and homework exercises with solution. Since there is nothing like teaching to truly learn a topic, the goal is for you to choose something of relevance/interest to you and thoroughly learn the statistical underpinnings. You can begin right away looking for a suitable article by choosing from articles you have already read (but not fully understood) for research, reading the primary literature, or talking to the instructor.

Computer Programming: Your bioinformatics skills list should ultimately include a few different programming languages (some of C, C++, Java, Python, Perl, R, and more). I do not enforce use of a particular language in this class, though we will sometimes rely on R for
statistical analysis. Although all writeups for this class should be independently prepared, you are invited to team up and share code for all programming assignments.

Topics:

- **Basic probability and statistics.**
  - **Week 1.** Review of basic probability and hypothesis testing.
  - **Week 2.** Estimation, likelihood, Bayesian framework. Simple sequence models.

- **Generative sequence models.**
  - **Models.** Markov chains, Hidden Markov models, stochastic grammars
  - **Applications.** biological patterns, motifs, genome annotation
  - **Week 3.** Markov chains.
  - **Week 4.** Motif finding. EM Algorithm.
  - **Week 5.** Hidden Markov Model.
  - **Week 6.** Finding similar sequences (BLAST and friends).

- **Week 7.** Midterm covering sequence models.

- **Comparative genomics.**
  - **Models.** extreme value distribution, Continuous Time Markov chain (CTMC), coalescent
  - **Applications:** sequence alignment, phylogenetic trees
  - **Week 8.** Evolutionary models.
  - **Week 9.** Phylogeny

- **High throughput sequence data**
  - **Models.** Count data (Poisson, Negative Binomial distribution)
  - **Applications.** variant calling, gene expression
  - **Week 10.** NGS. Base calling.
  - **Week 11.** NGS. SNP calling.
  - **Week 12.** RNA-seq. Differential expression.

- **Other statistical methods**
  - **Models.** Principal Components Analysis (PCA), machine learning (basic)
  - **Applications.** population structure, GWAS, metagenomics
  - **Week 13.** Exploratory methods.
  - **Week 14.** Multiple testing.

- **Week 15.** Project presentations.

- **Week 16.** Final exam.

**Course Policies**

**Academic Dishonesty:** The class will follow Iowa State University’s policy on academic dishonesty. Anyone suspected of academic dishonesty will be reported to the Dean of Students Office. [http://www.dso.iastate.edu/ja/academic/misconduct.html](http://www.dso.iastate.edu/ja/academic/misconduct.html).

**Disability Accommodation:** Iowa State University complies with the Americans with Disabilities Act and Sect 504 of the Rehabilitation Act. If you have a disability and anticipate needing accommodations in this course, please contact (instructor name) to set up a meeting within the first two weeks of the semester or as soon as you become aware of your need. Before meeting with (instructor name), you will need to obtain a SAAR form with recommendations for accommodations from the [Disability Resources Office](http://www.dso.iastate.edu/ja/academic/misconduct.html), located in Room 1076 on the main floor of the Student Services Building. Their telephone number is 515-294-7220 or email disabilityresources@iastate.edu. Retroactive requests for accommodations will not be honored.
Dead Week: This class follows the Iowa State University Dead Week policy as noted in section 10.6.4 of the Faculty Handbook [http://www.provost.iastate.edu/resources/faculty-handbook]

Harassment and Discrimination: Iowa State University strives to maintain our campus as a place of work and study for faculty, staff, and students that is free of all forms of prohibited discrimination and harassment based upon race, ethnicity, sex (including sexual assault), pregnancy, color, religion, national origin, physical or mental disability, age, marital status, sexual orientation, gender identity, genetic information, or status as a U.S. veteran. Any student who has concerns about such behavior should contact his/her instructor, Student Assistance at 515-294-1020 or email dso-sas@iastate.edu or the Office of Equal Opportunity and Compliance at 515-294-7612.

Religious Accommodation: If an academic or work requirement conflicts with your religious practices and/or observances, you may request reasonable accommodations. Your request must be in writing, and your instructor or supervisor will review the request. You or your instructor may also seek assistance from the Dean of Students Office or the Office of Equal Opportunity and Compliance.

Contact Information: If you are experiencing, or have experienced, a problem with any of the above issues, email academicissues@iastate.edu

January 13, 2015