

PHC6068 Biostatistical Computing

Instructor: Dr. Faming Liang, <http://users.phhp.ufl.edu/faliang/>

Time & Location: Monday 12:50–2:45pm (MAT 0151); Wednesday 1:55–2:45pm (ROG 0106)

Course Overview: This course is to introduce to students some key algorithms for statistical computing and their applications in biomedical sciences. The course will cover fundamental computational techniques for biostatistical data analysis as well statistical methods for random number generation, Monte Carlo integration, and stochastic optimization. Topics will include high dimensional data analysis, large-scale hypothesis testing, Markov chain Monte Carlo (MCMC), stochastic optimization, missing data analysis, and selected biological applications.

Prerequisites: One year graduate course of Probability and Statistical Inference

Course Objectives: Upon successful completion of the course, students should:

- Understand the role of computing techniques in biostatistical data analysis.
- Understand the nature of high dimensional data.
- Be aware of fundamental concepts of Markov chain Monte Carlo and stochastic optimization.
- Be aware of fundamental concepts of data reduction, model selection, and large-scale hypothesis testing.
- Be able to conduct Bayesian analysis for biostatistical data.
- Be able to analyze high dimensional data using recent statistical methods.

Topics to be covered:

1. Markov Chain Monte Carlo and Stochastic Optimization
 - (a) Basic algorithms: Metropolis-Hastings algorithm, Gibbs sampler.
 - (b) Advanced MCMC algorithms: parallel tempering, evolutionary Monte Carlo, stochastic approximation Monte Carlo, adaptive MCMC algorithms.
 - (c) Sequential Monte Carlo.
 - (d) Stochastic optimization: simulated annealing, genetic algorithms.
 - (e) The EM and stochastic approximation algorithms for missing data analysis.
2. High Dimensional Data Analysis
 - (a) Dimension reduction: principal component analysis, single value decomposition.
 - (b) Regularized regression: Lasso and related methods.

- (c) Bayesian variable selection, Bayesian model averaging.
- 3. Large-Scale Hypothesis Testing
 - (a) False discovery rate (FDR)
 - (b) Empirical Bayes methods
 - (c) Benjamini and Hochberg's FDR control methods.
- 4. Selected Biological Applications
 - (a) Hidden Markov models for sequence alignment.
 - (b) Next generation sequence data analysis.
 - (c) Phylogeny analysis.
 - (d) Genome-wide association studies.

Recommended texts/references:

1. Liu, J.S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer: New York.
2. Liang, F., Liu, C. and Carroll, R.J. (2010). *Advanced Markov chain Monte Carlo Methods: Learning from Past Samples*. Wiley.
3. Bühlmann, P. and van de Geer, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer.
4. Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. New York: Springer.
5. Efron, B. (2010). *Large-Scale Inference*. Cambridge: Cambridge University Press.
6. Ewens, W.J. and Grant, G.R. (2001). *Statistical Methods in Bioinformatics*. New York: Springer.
7. Gentleman, R., Carey, V., Huber, W., Irizarry, R.A., and Dudoit, S. (2005). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer.

Course Requirements/Evaluation/Grading

The assessment will include class participation, assignments, one course project, and one final exam. Class participation will include weekly attendance and participation in discussions. Students are responsible for all course material, including reading required materials prior to each class. Failure to complete assignments will result in a failing grade.

Class participation: 10% Assignments: 30% Course Project: 30% Final Exam: 30%

The grading scale for this course consists of the standard scale, including minus grades, below. The conversion factors for grade point values that are assigned to each grade are also included (in parentheses):

93% - 100% = A (4.00)

90% - 92% = A- (3.67)

87% - 89% = B+ (3.33)

83% - 86% = B (3.00)

80% - 82% = B- (2.67)

77% - 79% = C+ (2.33)

73% - 76% = C (2.00)

70% - 72% = C- (1.67)

67% - 69% = D+ (1.33)

63% - 66% = D (1.00)

60% - 62% = D- (0.67)

Below 60% = E (0.00)

Class Demeanor Expected by the Professor (late to class, cell phones): Students are expected to show up for class prepared and on time. Cell phones are to be silenced during class unless there is an emergency, in which case please inform the instructor.

Academic Integrity: Cheating, lying, misrepresentation, or plagiarism in any form is unacceptable and inexcusable behavior. Students are expected to act in accordance with the University of Florida policy on academic integrity (see Student Conduct Code, the Graduate Student Handbook or this web site for more details: www.dso.ufl.edu/judicial/procedures/academicguide.php).

*We, the members of the University of Florida community,
pledge to hold ourselves and our peers to the
highest standards of honesty and integrity.*

Policy Related to Class Attendance and Late or Missed Assignments: Attendance of all class sessions is required. Please see the instructor as early as possible regarding possible absences. All assignments need to be handed in on time. Grading will penalize late assignments. Missed assignments will receive a zero score. Personal issues with respect to class attendance or fulfillment of course requirements (assignments, final presentation, class discussion) will be handled on an individual basis.

Accommodations for Students with Disabilities: If you require classroom accommodation because of a disability, you must first register with the Dean of Students Office (www.dso.ufl.edu/drc). The Dean of Students Office will provide documentation to you, which you then give to the instructor when requesting accommodation. The College is committed to providing reasonable accommodations to assist students in their course work. If you have experienced problems in university classes

with writing, in-class exams, understanding or concentrating in class; please talk to us or access a learning or education testing resource at the University or in another professional setting.

Counseling and Student Health: If you find yourself facing personal problems affecting your course work, you are encouraged to talk with an instructor and to seek confidential assistance at the University of Florida Counseling Center, 352-392-1575. Visit their web site for more information: www.counseling.ufl.edu

The Student Health Care Center has three locations: on Fletcher Drive, at Shands, and at Corry Village. It offers a variety of clinical services, including primary care, women's health care, immunizations, mental health care, and pharmacy services. For more information, contact the clinic at 352-392-1161 or check out the web site at: shcc.ufl.edu

Crisis intervention is available 24/7 from the Alachua County Crisis Center: (352) 264-6789.

BUT - Do not wait until you reach a crisis to come in and talk with us. We have helped many students through stressful situations impacting their academic performance. You are not alone so do not be afraid to ask for assistance.