140.778 ADVANCED STATISTICAL COMPUTING  
(Biostatistics – 2nd term, 3 units)

COURSE SYLLABUS  
Instructor: Hongkai Ji

Contact  
Office: 615 N. Wolfe St., Rm. E3638  
Office hours: Thursday 3:00-4:00pm  
Office Phone: 410-955-3517  
Email: hji@jhsph.edu  
Course page: http://www.biostat.jhsph.edu/~hji/courses/computing/

Class times  
Tuesday 8:30 - 9:50am  
Thursday 8:30 - 9:50am

Location  
Wolfe W4013

Description  
This course covers the theory and application of common algorithms used in statistical computing. Topics include root finding, optimization, numerical integration, Monte Carlo, Markov chain Monte Carlo, stochastic optimization and bootstrapping. Some specific algorithms discussed include: Newton-Raphson, EM, Metropolis-Hastings algorithm, Gibbs sampling, simulated annealing, Gaussian quadrature, Romberg integration, etc. Applications of these algorithms to real research problems will be discussed.

Course Learning Objectives  
Upon successfully completing this course, students will be able to: 1) describe common deterministic statistical algorithms, such as root finding, numerical integration methods, Newton-Raphson, quasi-Newton methods, EM, MM; 2) describe common stochastic algorithms used in statistics, such as Monte Carlo methods, Markov Chain Monte Carlo, stochastic optimization, Gibbs sampling, Metropolis-Hastings method; 3) understand mathematical properties of common statistical algorithms; 4) implement statistical algorithms using a high-level statistical programming language.

Prerequisites  
Prior programming experience; at least one year of doctoral-level statistics/biostatistics theory and methods courses; 140.776

Texts & References  
Numerical Analysis for Statisticians, Kenneth Lange, Springer  
Monte Carlo Strategies in Scientific Computing, Jun S. Liu, Springer
Grading policy
Attendance: 10%
Homework: 60%
Final Project: 30%

Schedule
Oct 20 (Thu): Introduction
Oct 25 (Tue): Solution of nonlinear equations: bisection, functional iteration, Newton’s method
Oct 27 (Thu): Convergence rate, Optimization (homework 1)
Nov 1 (Tue): Newton-Raphson, Scoring
Nov 3 (Thu): EM algorithm, Applications of EM
Nov 8 (Tue): Convergence rate of EM, Model-based Clustering
Nov 10 (Thu): Mixtures of Factor Analyzers, MM algorithm (homework 1 due; homework 2)
Nov 15 (Tue): Random number generation
Nov 17 (Thu): Markov chains
Nov 22 (Tue): Markov Chain Monte Carlo (MCMC), Metropolis-Hasting
Nov 24 (Thu): Thanksgiving, no class
Nov 29 (Tue): Gibbs sampler, MCMC in DNA motif discovery
Dec 1 (Thu): MCMC in DNA motif discovery (homework 2 due, homework 3)
Dec 6 (Tue): Marginalization, General conditional sampling
Dec 8 (Thu): Importance sampling
Dec 13 (Tue): Variance reduction, Gaussian quadrature
Dec 15 (Thu): Romberg integration (homework 3 due)

Final Project Due Date
Dec 30 (Thu)