ROLLINS SCHOOL OF PUBLIC HEALTH
EMORY

DEPARTMENT: **BIOSTATISTICS AND BIOINFORMATICS**

COURSE NUMBER: **732** **SECTION NUMBER:** **1** **SEMESTER: Spring**

CREDIT HOURS: 2

COURSE TITLE: Advanced Numerical Methods

INSTRUCTOR NAME Mary E. Kelley, PhD

INSTRUCTOR CONTACT INFORMATION

EMAIL: mekelle@emory.edu

PHONE: (404) 712-0804

SCHOOL ADDRESS OR MAILBOX LOCATION:

OFFICE HOURS Th 2:30 – 4:00

BRIEF COURSE DESCRIPTION

The course covers topics in traditional numerical analysis specifically relevant to statistical estimation and inference. The topics covered include numerical linear algebra, the root finding problem (maximum likelihood) methods, and both frequentist and Bayesian techniques for further exploration of the likelihood. If there is time, numerical integration and curve fitting may also be presented. Prerequisites include BIOS 710, 711: BIOS 707,711 may be taken concurrently.

LIST SCHOOL LEVEL, DEPARTMENT, AND/ OR PROGRAM COMPETENCIES

School:

- Use analytic reasoning and quantitative methods to address questions in public health and population-based research

Department:

- Conduct independent research
- Assist in the development of new statistical methods as needed to address public health or medical problems
- Apply new and existing statistical theory and methods as needed to address public health problems

LIST LEARNING OBJECTIVES ASSOCIATED WITH THE COMPETENCIES

- Provide doctoral students the basics of the numerical techniques which will be employed in their thesis work
- to understand the methods and algorithms which are employed by the accessible routines in common statistical packages such as SAS, R, and Splus, and to be able to construct similar algorithms for the development of new methods
- to be able to critically evaluate newly developed techniques from a numerical and computing perspective and use them to perform analyses of research data

EVALUATION

70% of the grade will be based on homework assignments. Students will also present a paper illustrating a relevant numerical technique to the class; the presentation will be 30% of the grade. Papers will be chosen by the instructor.

ACADEMIC HONOR CODE
The RSPH requires that all material submitted by a student in fulfilling his or her academic course of study must be the original work of the student.

Syllabus: Advanced Numerical Methods

Required textbooks:

- Tanner, MA. Tools for Statistical Inference. Methods for the Exploration of Posterior Distributions and Likelihood Functions, 3rd edition. Springer, New York, 1996. ISBN: 0387946888
- Lange, K. Numerical Analysis for Statisticians, 2nd edition. Springer, New York, 2010. ISBN: 1441959440

Supplemental readings:

- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. JRSS-B, 39:1-38, 1977.

1) Numerical linear algebra – solving linear systems (Generalized (weighted) least squares)
   a) Direct methods – Lange chapter 7
      i) (P) LU factorization
      ii) Cholesky decomposition
      iii) QR factorization (Gram-Schmidt orthogonalization)
   b) Solution stability (vector/matrix norms) Lange 6.1-6.4,6.6
   c) Iterative methods – convergence (ex. Jacobi iteration Lange 6.5)
   d) The eigenvalue problem (multivariate, data reduction) Lange 8.1,8.2
      (1) QR method
      (2) SVD, PCA
2) Solving nonlinear systems – root finding
   a) Bisection (Lange 5.2)
   b) Newton's method (NR algorithm) and Fisher scoring (Lange 5.4, 14.1 – 14.4; Tanner 2.3)
   c) IRLS (Lange 14.4)
   d) Quasi-Newton methods (Lange 14.9)
   e) EM algorithm and variations
      i) general method, comparison to NR (Lange 13.1-13.3, 13.6, Tanner 4.1-4.2)
      ii) special cases
         (1) exponential family (DLR, Tanner 4.3)
         (2) Bayesian (DLR, Lange 13.5)
         (3) Missing data/imputation (13.4)
         (4) Mixture models (DLR, 13.7)
      iii) ECM, ECME (Tanner 4.7)
      iv) Standard errors and EM (Tanner 4.4)
   f) Convergence and ascent properties of fixed point (root finding or optimization) methods (Lange 13.3. Ch 15)
   g) Acceleration methods (Tanner 4.6)
3) Numerical methods for marginalization (approximating integrals) and sampling from composite distributions (inference)
   a) Laplace approximation (Tanner 3.2)
   b) Monte Carlo methods (Tanner 3.3)
      i) composition
      ii) importance sampling-rejection/acceptance (Lange 22.5,23.2)

c) MCMC methods (Tanner 4.5, Ch 5, Ch6; Lange Ch 26)
    i) Data augmentation
    ii) Gibbs sampling
    iii) Metropolis algorithm
d) Numerical integration and quadrature (Tanner 3.1, Lange 18.6)

4) *Optional module*: polynomial fitting and splines
5) Comparison of missing data solutions (EM, MCMC)?