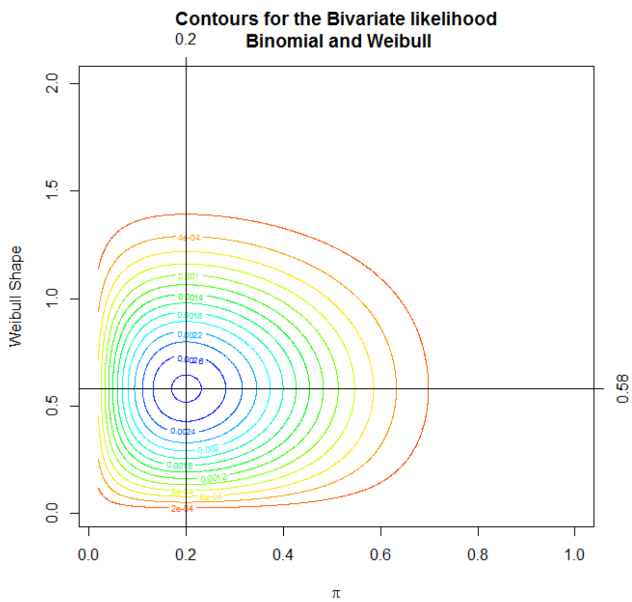
MATH 5793/4793 2015: Advanced Applied Statistics Syllabus

Introduction to the course

This course comprises an introduction to Multivariate Statistics.
The course does not assume a depth of statistical knowledge but students without a background in univariate statistics (like MATH 4753) should be attentive when analogies to univariate statistics are made. Most ideas will be taught from the ground up assuming little background in statistics.
A good background in linear algebra will benefit students.
I will give an in depth summary of R (Statistical Software) and students will have to become proficient at making functions.

Why do it?

Most data we collect and process in science is multivariate. It behoves the scientist to learn good methods to both graphically and analytically summarize complex data sets.
Further – how to perform tests based on model assumptions will be important.



This course will equip you with the theory behind a number of statistical techniques (see below) and train you in how to actually perform the analyses in R.
The following is a summary of the syllabus likely covered.

Syllabus

1. Introduction to the R software
    a. Making functions
    b. Plotting
    c. Returning statistical output
    d. Libraries
    e. Editors
2. Organization of data
    a. Arrays
    b. Descriptive statistics
    c. Graphical techniques
        i. Chernoff faces
        ii. stars
        iii. pairs plots
        iv. 3-d plots
    d. Statistical distance
3. Matrix algebra and random vectors
    a. Vectors
    b. Matrices
    c. Positive definite matrices
    d. Square root matrix
    e. Random vectors and matrices
    f. Mean vectors and covariance matrices
        i. Partitioning the covariance matrix
        ii. Mean vector and Covariance Matrix for linear combinations of random variables
        iii. Partitioning the sample mean vector and covariance matrix
    g. Matrix inequalities and maximization
4. Sample Geometry and random sampling
    a. Geometry of the sample
    b. Expected values of sample mean and covariance matrix
    c. Generalized variance
        i. GV=0
        ii. |R| and geometrical interpretation
        iii. Trace of the diagonal
    d. Sample mean, covariance and correlation as matrix operations
    e. Sample values of linear combinations of variables
5. The multivariate Normal.
    a. The multivariate normal density and its properties
    b. Sampling from a multivariate distribution and maximum likelihood
        i. The multivariate normal likelihood
        ii. Maximum likelihood estimation of $\mu\ and\ \Sigma$
        iii. Sufficient statistics
    c. The sampling distribution of $\bar{X}$ and S
    d. Assessing assumption of normality

    e. Detecting outliers
    f. Transformations to Normality using BOX COX
6. Principal components
    a. Population principal components
        i. Standardized variables
        ii. Covariance matrices with special structures
    b. Summarizing Sample variation by PCA
        i. Number of principal components
        ii. Interpretation of principal components
        iii. Standardizing the sample pc's
    c. Graphing them
    d. Large sample inferences
    e. Quality control
        i. Ellipses
        ii. T^2 plots
7. Factor analysis
    a. Orthogonal factor model
    b. Methods of estimation
        i. Principal components
        ii. Modified approach
        iii. Max. Lik. Method
        iv. Large sample test for common factors
    c. Factor rotation
    d. Factor scores
    e. Strategies for factor analysis
    f. Structural equation models