Date: 20 April 2019

Syllabus for PhD Applied Statistics

Linear Algebra	Vector spaces, orthogonality of vectors, Systems of linear equations, Characteristic roots and vectors, spectral decomposition, Moore-Penrose and generalized inverses and their properties. Solution of matrix equations. Sufficient – Condition for the existence of homogenous and non-homogenous linear equation, Real quadratic forms.
Probability distribution theory	Elements of probability, Probability basics, Discrete (integer valued) random variable, Real and vector-valued random variables, distribution functions (d.f.), Properties of Expectations, Discrete distributions, Continuous distributions: Special distributions: Cauchy, Laplace, t and F distributions including their non-central versions. Classes of sets, fields and sigma-fields, limit of sequences of subsets, sigma-field generated by a class of subsets, Borel fields, Independence of two events and n (> 2) events, sequence of independent events, independent classes of events, Convergence of a sequence of r.v.s., Various types of convergence and their interrelationships, Cramer's theorem (Slutsky's theorem), monotone convergence theorem and dominated convergence theorem, Unlimited Sequences of Bernoulli Trials, Recurrent Events, Random Walk and Ruin Problems.
Sampling Theory	Concept of sampling, need for sampling, population and sample, sampling unit and sampling frame, Types of Population, Basic properties of population, sample survey and census, Principal steps in a Sample survey, Sampling and non-sampling error, Simple Random Sampling, Estimation of Population mean and proportion and their sampling variances, Determination of sample size, Stratified sampling, Systematic sampling, Cluster sampling, Multi-stage sampling, Large scale sample surveys, Sources of Non sampling errors and methods of controlling them, NSSO, NFHS and CSO and their function.
Linear Models	Hypothesis testing in regression, Confidence intervals in regression, Prediction of new observations, Coefficient of determination Simple linear regression, assumptions, inference related to regression parameters, standard error of prediction, tests on intercepts and slopes, extrapolation, Diagnostic checks and correction: graphical techniques, tests for normality, uncorrelatedness, homoscedasticity, lack-of-fit testing, polynomial regression, Multiple regression: standard Gauss Markov setup, least squares (LS) estimation, variance-covariance of LS estimators, estimation of error variance, Regression analysis with correlated observations, Regularized regression models. Generalized Linea Model, Neural network approach to regression.

Statistical Computing	Random number generation, Tests of randomness, digit frequency test and serial correlation, selection of a random number generator, Optimization methods, direct search, grid search, interpolator search, gradient search. Newton-Raphson method, Muller's method, Aitken's extrapolation, Numerical integration, Monte Carlo Methods, Finding zeros of function: interpolation, Newton-Raphson method, Secant method etc.
Statistical Inference	Sufficiency, sufficient partition, Neyman factorization theorem, minimal sufficiency, Consistency of estimators and its properties including invariance, completeness, Ancillarity and Basu's theorem, Exponential family of distributions, Unbiased Estimator, Estimability of parametric functions, Cramer-Rao inequality, Uniformly minimum variance unbiased estimators, Rao-Blackwell and Lehmann Scheffe theorems, Point estimation and interval estimation, Asymptotic distribution of log-likelihood ratio, UMP Unbiased tests, Non parametric tests, Sign test, Wilcoxon sign ranked test, Mann-Whitney U test, Kruskal-Wallis test.
Design of Experiments	Principles of Experimental Design, Connectedness, Balanced and Orthogonality, ANCOVA, Factorial Design, 2^k Factorial Design, 3^k Factorial Design, Full and fractional replicates, confounding, Fractional Factorial Experiments, and Fractional Factorial, Response Surface Methodology, Taguchi Methods: Noise Factor, Loss Function, Signal to Noise Ratio, Orthogonal Arrays, Triangular Tables, Linear graphs, Inner and outer arrays, Random Effect Models.
Multivariate Statistical Analysis	Multivariate Normal Distribution, Quadratics form and its distributions, Cochran's theorem., Wishart distribution, derivation and properties, additivity, determinant of W, distribution of multiple and partial correlation, Hoteling's T^2, Inference on Multivariate Normal distribution, MLE of mean vector, Variance-Covariance Matrix. Principal Component Analysis(PCA), Factor Analysis(FA), Discriminant Analysis, Classification Problem, Introducing LDF and FDF as supervised methods, Likelihood ratio test.
Computer Intensive Statistical Methods	Resampling Techniques, Missing Values and Imputations Techniques, EM Algorithm and Applications, Smoothing techniques, Bayesian computing, Markov Chain Monte Carlo. Simulation using MCMC, Particle filtering, MCMC methods for missing values
Statistical Learning and Data Mining	Supervised Learning: K - nearest neighbourhood algorithm, Decision trees, Neural networks, Naïve Bayes and Bayesian networks. Kernel methods, Support Vector Machines, Model evaluation techniques, Cost-Benefit analysis using data driven costs. Unsupervised Learning: Hierarchical and k-means clustering, Kohonen networks, BIRCH clustering, Measuring cluster goodness.

Stochastic Processes	Markov chain, transition probabilities, Chapman-Kolmogorov equations, evaluation of higher step transition probabilities, stationary distribution, Classification of states, periodicity of a Markov chain, Branching process, Extinction probability, Introduction to birth process, birth and death process. Linear birth and death process. Growth model with immigration and related results, Poisson process, two definitions and their equivalence. Distribution of inter arrival times, conditional joint distribution of inter arrival times. Compound Poisson process, Renewal process and Weiner process.
Statistical Simulation	Stochastic simulations: generating random variables, simulating normal, gamma and beta random variables, Comparison of algorithms to generate random variables, Simulating multivariate distributions, Variance reduction techniques, Simulating a non-homogeneous Poisson process, Optimization using Monte Carlo methods simulated annealing for optimization. Solving differential equations by Monte Carlo methods
Methods for high dimensional data and machine learning	Supervised classification models, decision trees/ classification tree, recursive partitioning, CART algorithm, Support vector machines, Artificial Neural Networks, Learning models in ANN, Unsupervised Classification Models, K-means clustering algorithms, K-mediods algorithm, determining value of K
Optimization Techniques	Linear Programming, Integer linear programming, Fractional cut method- all integer and mixed integer linear programming problem, branch and bound method, dynamic programming, sensitivity. Bellman's optimality principle, Review of transportation and assignment problems, Nonlinear programming, Networking model, Network scheduling by PERT/CPM Techniques. Resource Analysis in network scheduling