University of Pune

Two Year M. A. / M. Sc. Degree Program in Statistics

Revised Syllabi of Second Year M. A. / M. Sc. in Statistics (Credit System)

(To be implemented in the Department of Statistics, University of Pune)

(With effect from Academic Year 2014-2015)

Submitted by

Prof. S. R. Deshmukh
Chairperson,
Board of Studies in Statistics

M.A. / M. Sc Statistics program has semester pattern and credit system with variable credits. The program consists of 100 credits. Credits of a course are specified against the title of the course. A course with (**T)** indicates that it is a theory course whereas a course with (**P)** indicates that it is a practical course. Some of the practical courses are linked with a theory course and in such a case, both the courses will have the same course number with T and P, indicating a theory and a practical course respectively.

The program has **pre-requisites system**. Under this system, a student has to clear pre-requisite courses to take up some courses in following semesters. Unless student passes the pre-requisites of a course, (s) he will not be allowed to enroll in the said theory courses. Pre-requisites of a course are indicated in curly brackets against the course. The Department committee may change the pre-requisites of a course.

The revised syllabus for the first year has been implemented from the academic year 2013-14. In the first year all courses are compulsory. In semesters III and IV some courses are compulsory and others are elective. Following is a structure of the two year M.Sc. (Statistics) program.

Compulsory courses: Compulsory courses in semester I and semester II are listed below.

Semester I

ST 1	Mathematical Analysis (4)
ST 2 (T)	Linear Algebra (3)
ST 2 (P)	Linear Algebra (2)
ST 3	Univariate Probability Distributions (2)
ST 4	Multivariate Probability Distributions (4)
ST 5 (T)	Elements of Statistical Computing (4)
ST 5 (P)	Elements of Statistical Computing (3)
ST 6	Calculus (4)

Total credits 26

Semester II

ST 7	Probability Theory (4) {ST 1, ST 6}
ST 8	Limit Theorems in Probability (2) {ST 1, ST 6}
ST 9 (T)	Multivariate Analysis (4) { ST 2 (T), ST 3, ST 4, ST 6}
ST 9 (P)	Multivariate Analysis (2) { ST 2 (T), ST 3, ST 4, ST 6}
ST 10	Point & Interval Estimation (3) {ST 3, ST 4}
ST 11	Testing of Hypotheses (3) {ST 3, ST 4}
ST 12 (T)	Regression Analysis (4) {ST 2 (T), ST 3, ST 4}
ST 12 (P)	Regression Analysis (2) {ST 2 (T), ST 3, ST 4}

Total credits 24

Semester III

ST 13 Stochastic Processes (4)
ST 14 (T) Asymptotic Inference (4) {ST 7, ST 8, ST 10, ST 11}
ST 14 (P) Asymptotic Inference (2) {ST 7, ST 8, ST 10, ST 11}
ST 15 (T) Design of Experiments and Analysis of Variance (4) {ST 12 (T)}
ST 15 (P) Design of Experiments and Analysis of Variance (2) {ST 12 (T)}
Module specific/Elective course (4)
Module specific/Elective course (4)

Total credits 24

Semester IV

ST 16 (T) Sampling Methods (4)
ST 16 (P) Sampling Methods (2)
ST 17 Bayesian Inference (4) {ST 10, ST 11}
ST 18 (P) Project (4) {ST 9 (T), ST 12 (T)}
Module specific/Elective course (4)
Module specific/Elective course (4)
Module specific/Elective course (4)

Total credits 26

Detailed syllabi, along with the list of recommended books, of the core courses for the second year of M.Sc. are given below. Here L stands for Lectures and T for Tutorials. A tutorial session is equivalent to one Lecture wherein the students are expected to solve problems from pre-circulated set of problems on specified topics.

Courses in Semester III

ST 13 Stochastic Processes (4 Credits)

Unit 1

Notion of stochastic processes, Markov chain, one step transition probabilities, Chapman-Kolmogorov equations, evaluation of higher step transition probabilities, classification of states, periodicity of a Markov chain, concept of closed class, minimal closed class, stationary distribution. Some examples such as gamblers ruin problem and one dimensional random walk. Concept of absorption probabilities. Use of these to compute probability of winning the game by a gambler having initial capital 'a'.

(12L+3T)

Unit 2

Branching process, classification of states, identification of criticality parameter. extinction probability, relationship between criticality parameter and extinction

probability of the process. Expression for mean and variance of the process. Method to compute the extinction probability. Some epidemiological applications.

Introduction to Markov chain in continuous time, concept of intensity rate, relationship between intensity matrix and transition probability matrix. Kolmogorov's forward and backward equations.

(12L+3T)

Unit 3

Introduction to birth process, birth and death process. Linear birth and death process. Growth model with immigration and related results. Expression for mean and variance of a birth process and, birth and death process. Simple applications of these processes.

(12L+3T)

Unit 4

Poisson process, two definitions and their equivalence. Distribution of inter arrival times, conditional joint distribution of inter arrival times. Compound Poisson process. Some applications.

Introduction to renewal process, relationship with Poisson process, key and elementary renewal theorems associated with renewal processes. Some applications. Introduction to Weiner process and its simple properties.

Books Recommended

1. Bhat B.R. (2000). *Stochastic Models: Analysis and Applications*, New Age International.
2. Medhi, J. (1982) *Stochastic Processes*, Wiley Eastern.
3. Pinsky M. A. and Karlin, S. (2010). *An Introduction to Stochastic Modeling*, 4th Edn. Academic Press.
4. Ross, S. (2007). *Introduction to Probability Models*, 9th Edn., Academic Press.

Additional Books for Reference

1. Feller, W. (1972). *An Introduction to Probability Theory and its Applications*, Vol. 1, Wiley Eastern.
2. Hoel, P.G. Port, S.C. & Stone, C.J. (1972). *Introduction to Stochastic Processes*, Houghton Mifflin.
3. Karlin, S & Taylor, H.M. (1975). *A First Course in Stochastic Processes* (Second. Edition), Academic Press.
4. Serfozo, R. (2009). *Basics of Applied Stochastic Processes*, Springer.

ST 14 (T) Asymptotic Inference (4 Credits)

Unit 1

Consistent estimator for real and vector valued parameter, joint consistency and marginal consistency and equivalence between them, invariance property under

continuous transformation of consistent estimators for real and vector valued parameters

Methods for generating consistent estimators for real and vector valued parameters:

(i) method of moments: generating consistent estimator using weak law of large numbers

(ii) method of percentiles: consistency of sample percentiles for population percentiles

Comparison of consistent estimators based on (i) mean squared error, (ii) minimum sample size required by the estimator to attain certain level of accuracy

(12 L + 3 T)

Unit 2

Consistent and asymptotically normal (CAN) estimators: Definition of CAN estimator for real and vector valued parameter, invariance property of CAN estimator (delta method), and variance stabilizing transformation

Generating CAN estimators for real and vector valued parameter:

(i) method of moments: generating CAN estimator using central limit theorem

(ii) method of percentiles: generating CAN estimator using asymptotic distribution of sample percentiles

Comparison of CAN estimators based on asymptotic variance **(12 L + 3 T)**

Unit 3

Maximum Likelihood Estimation:

(i) MLE in case of restricted parameter space, MLE of parametric function, Inconsistent MLEs, MLEs in irregular cases.

(ii) Asymptotic distribution of MLE in special class of distributions: Cramer regularity conditions, Cramer- Huzurbazar theorem, Extension to vector-valued parameters, exponential family with natural parameters (canonical set up), its extension to the (usual) exponential family set up and multinomial distribution

(iii) Super efficient estimators, Hodges-Lecam shrinkage technique

(12 L + 3 T)

Unit 4

Asymptotic theory of tests of hypotheses: Tests based on MLEs. Likelihood ratio tests, asymptotic distribution of log likelihood ratio, Bartlett correction, Wald test, Score test, Pearson's chi-square test and LR test, Consistent Test

Comparison of tests: Asymptotic relative efficiency of tests, Pitman's asymptotic relative efficiency.

Asymptotic confidence intervals: Construction and examples

Applications: Applications to categorical data analysis, tests for independence for three dimensional contingency tables.

(12 L + 3 T)

Books Recommended

1. Ferguson, T.S. (1996), *A Course in Large Sample Theory*, Chapman & Hall, London, Chapters 1, 2, 4.
2. Gupta Anirban Das (2008), *Asymptotic Theory of Statistics and Probability*, Springer, New York. Chapters 1, 4, 7, 13, 16, 21, 22, 27.
3. Kale, B. K. (1999), *A First Course in Parametric Inference*, Narosa, Chapters 5, 6, 7, 9.
4. Lehmann, E. L. and Casella, G. (1998)*, Theory of Point Estimation,* Springer*,* New York, Chapter 6.
5. Rao, C. R. (1995). *Linear Statistical Inference and its Applications*, Wiley, New York, Chapters 5c-5g, 6a-6e.

ST 14 (P) Asymptotic Inference (2 Credits)

R will be used for the practical.

Unit 1

1. Verification of consistency of the estimators
2. Verification asymptotic normality of the estimators
3. Comparing methods of estimation, MSE and sample size considerations
4. Maximum likelihood estimation under various set up (includes censoring)

Unit 2

1. Analysis of higher dimensional contingency tables.
2. Verification of consistency of test procedure
3. Likelihood ratio tests
4. Verification of asymptotic distribution of likelihood ratio test statistic in case of Cramer family

ST – 15(T): Design and Analysis of Experiments (4 Credits)

Unit 1

Review of randomization, replication and local control. Analysis of one way classification model. Analysis of two way classification model with equal number of observations per cell with and without interactions. Analysis of two way classification model with unequal number of observations per cell without interactions. Notion of connectedness, balance and orthogonality. Analysis of

BIBD. Analysis of covariance in one way and two way classification models. Testing of hypotheses for estimable parametric functions.

(12L+3T)

Unit 2

Analysis of 2^k full factorial experiments: diagrammatic presentation of main effects and first and second order interactions, model, analysis of single as well as more than one replicates, using ANOVA. Total confounding of 2^k design in 2^p blocks, $p \geq 2$. Partial confounding in 2^p blocks, p = 2, 3. Fractional factorial experiments. Resolution of a design, (III, IV & V), aberration of a design.

(12L+3T)

Unit 3

Analysis of 3^2 designs: contrasts for linear and quadratic effects, analysis of 3^2 design, confounding and fractional experiments in 3^2 design. Response surface methodology (RSM): linear and quadratic model, stationary point, central composite designs(CCD), ridge systems, multiple responses, concept of rotatable designs, Box-Behnken design, optimality of designs, simplex lattice designs, simplex centroid designs. **(12+3T)**

Unit 4

Taguchi methods: concept of noise factors, concept of loss function, S/N ratio, orthogonal arrays, triangular tables, linear graphs, inner and outer arrays. Random effect models and mixed models.

(12L+3T)

Books Recommended:

1. Dean, A. and Voss, D. (1999). *Design and Analysis of Experiments*, Springer.
2. George E. P. Box, Draper N.R. (1987). *Empirical Model-Building and Response Surfaces*, Wiley.
3. Hicks, C.R., Kenneth V. and Turner, Jr. (1999). *Fundamental Concepts in the Design of Experiments*, Oxford University Press.
4. Kshirsagar A.M. (1983). *Linear Models*, Marcel Dekker
5. Montgomery, D.C. (2001). *Design and Analysis of Experiments*, Wiley.
6. Ogawa, J. (1974). *Statistical Theory of the Analysis of Experimental Design*, Marcel Dekker.
7. Phadke, M.S. (1989). Quality Engineering using Robust Design, Prentice Hall, Englewood Cliffs, New Jersey
8. Wu, C.F. Jeff and Hamada M. (2000). Experiments: Planning ,Analysis and Parameter Design Optimization, John Wiley and Sons

ST 15 (P): Design and Analysis of experiments (2 Credits)

Unit 1

1. One way classification. Multiple comparison tests

2. Two way classification with equal number of observations per cell (Model with interaction). Two way classification with unequal number of observations per cell (Model without interaction)
3. Analysis of LSD and BIBD.
4. Analysis of covariance in one way and two way model.
5. 2^k factorial experiments, Analysis of single replicate of 2^k.
6. Total and partial confounding in 2^k factorial experiments.

Unit 2

7. Analysis of 2^k fractional factorial experiments
8. Analysis of 3^2 factorial experiments
9. Random effect and mixed models
10. Analysis of first and second order response surface model. Central composite design. Contour and surface plots, Box-Behnken design
11. Small composite designs, optimality of designs, simplex lattice designs, simplex centroid designs.
12. Taguchi methods: S/N ratio, orthogonal arrays, triangular tables, linear graphs, inner and outer arrays.

1) Author: C.F. Jeff Wu, Michael Hamada: (2000)
Name: Experiments: Planning ,Analysis and Parameter Design Optimization.
Publisher: John Wiley and Sons
2) Author: Madhav S. Phadke (1989)
Name: Quality Engineering using Robust Design
Publisher:

Courses in Semester IV

ST 16 (T): Sampling Methods (4 credits)

Unit 1

Review of basic methods of sample selection, simple random sampling with replacement (SRSWR), simple random sampling without replacement (SRSWOR). Use of supplementary information for estimation, ratio and regression estimators with their properties and generalizations, systematic sampling. **(12L+3T)**

Unit2

Review of stratification: Allocation problems and estimation problems.
Formation of strata and number of strata. method of collapsed strata, Post stratification, probability proportional sampling with and without replacement, estimation problems, Hansen-Horwitz estimator and its properties, Horwitz-Thompson estimator and its properties, Midzuno-Sen method.

(12L+3T)

Unit 3

Cluster sampling, multistage-sampling. Double sampling procedures and their ratio and regression estimators, stratification estimator. Multiphase sampling. **(12L+3T)**

Unit 4

Non-sampling errors, response and non-response errors, Hansen and Hurwitz's model and their treatments, methods of imputation, randomized response, Warner's model, Franklin's model. Jackknife technique.

(12L+3T)

Books Recommended

1. Cochran, W.G. (1984). *Sampling Techniques*, Wiley.
2. Des Raj and Chandhok, P. (1998). *Sample Survey Theory*, Narosa.
3. Murthy, M.N. (1977). *Sampling Theory and Methods*, Statistical Publishing Society
4. Okafore, C (2002). *Sample survey Theory with Applications*, Snaap Press Ltd.
5. Singh, D. and Chaudhary F.S (1986).*Theory and Analysis of Sample Survey Designs*, Wiley Eastern Limited.
6. Singh, S. (2003). *Advance Sampling Theory and Applications* (Volume I and II), Kluwer Academic Publishers.
7. Sukhatme P.V, Suktatme, B.V.,Sukhatme S. and Asok C. (1984). *Sampling Theory of Surveys with Applications*, Indian Society for Agricultural Statistics, New Delhi.

ST 16 (P): Sampling Methods (2 credits)

Unit 1

1. SRSWOR, SRSWR,
2. Stratified random sampling, various kinds of allocation, post stratification, use of auxiliary information.
3. Ratio and regression methods of estimation.
4. Probability proportional to sizes sampling design.
5. Double sampling.

Unit 2

6. Two stage sampling.
7. Systematic sampling.
8. Cluster sampling
9. Randomized response technique.
10. Two practicals of consolidated nature each one of which would use theory of one or more of the above topics or a sampling problem under a real set up (including estimation)

ST-17: Bayesian Inference (4 credits)

Unit 1

Basics of minimaxity. Subjective and frequentist probability, Bayesian inference, prior distributions, posterior distribution, loss function, principle of minimum expected posterior loss, quadratic and other common loss functions, advantages of being a Bayesian. Improper priors, common problems of Bayesian inference, Point estimators. Bayesian HPD confidence intervals, testing, credible intervals, prediction of a future observation,

(12L + 3 T)

Unit 2

Bayesian analysis with subjective prior, robustness and sensitivity, classes of priors, conjugate class, neighborhood class, density ratio class, different methods of construction of objective priors: Jeffrey's prior, probability matching prior, conjugate priors and mixtures, posterior robustness: measures and techniques

(12L + 3 T)

Unit 3

Model selection and hypothesis testing based on objective probabilities and Bayes factors

Large sample methods**:** Limit of posterior distribution, consistency of posterior distribution, asymptotic normality of posterior distribution

(12L + 3 T)

Unit 4

Bayesian Computations: Analytic approximation, E-M Algorithm, Monte Carlo sampling, Markov Chain Monte Carlo Methods, Metropolis-Hastings Algorithm, Gibbs sampling, examples, convergence issues

(12L + 3 T)

Books Recommended

1. Bolstad, W. M. (2007). *Introduction to Bayesian Statistics*, 2nd Edn. Wiley.
2. Christensen R, Johnson, W., Branscum A. and Hanson T. E. (2011). *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*, Chapman & Hall.
3. Congdon, P. (2006). *Bayesian Statistical Modeling*, Wiley
4. Ghosh, J. K., Delampady M. and T.Samantha (2006). *An Introduction to Bayesian Analysis: Theory & Methods*, Springer.
5. Jim, A. (2009). *Bayesian Computation with R*, 2nd Edn, Springer.
6. Lee, P. M. (2004). *Bayesian Statistics: An Introduction*, Hodder Arnold.
7. Rao. C.R. and Day. D. (2006). *Bayesian Thinking, Modeling & Computation*, *Handbook of Statistics*, Vol. 25. Elsevier

Elective Courses

The syllabi of some module specific and elective courses, which are being taught during the last few years is given below. These module specific and elective courses will be taught in the coming years, depending on the availability of the faculty. All these courses are of 4 credits.

Module Specific and Elective Courses

ST P1: Measure Theory and Probability

Unit 1

Ring, σ-ring, Measure, measure space, Caratheodory Extension theorem, Lebesgue measure. **(12L + 3T)**

Unit 2

Integral of a measurable function with respect to a measure, its properties. Hahn – Jordan decomposition, Lebesgue decomposition, Radon – Nikodym derivative. product measure, Fubini's theorem. **(12L + 3T)**

Unit 3

Convergence in measure, almost everywhere convergence, Kolmogorov inequality. Kolmogorov three series criterion and strong law of large numbers. Introduction to weak convergence. **(12L + 3T)**

Unit 4

Conditional probability and conditional expectations, their simple properties. Martingales, martingale convergence theorems (SLLN, CLT). **(12L + 3T)**

Books Recommended

1. Ash, R.B. (1972). *Real Analysis and Probability.* Academic Press
2. Athreya, K.B. and Lahiri, S.N. (2006). *Measure Theory and Probability Theory*. Springer.
3. Billingsley, P. (1986). *Probability and Measure.* John Wiley
4. Taylor, J. C. (1997). *Introduction to Measure and Probability.* Springer.
5. Williams, D. (1991). *Probability with Martingales*. Cambridge University Press.

Additional Books for Reference

1. Ash, R. B. (2000). *Probability & Measure Theory*. Academic Press.
2. Dudley, R. M. (2004). *Real Analysis and Probability*. Cambridge University Press.
3. Halmos, P. R. (1978). *Measure Theory*, Springer.

ST P2: Advanced Stochastic Processes

Unit 1

Markov Chains: Taboo probabilities and ratio limit theorems. Invariant measures: [1]. ch. 11, Sections 1-4. **(12L + 3T)**

Unit 2

Markov Sequences: Definitions, transition densities, stationary distribution, normal Markov sequences [2] Ch. VI Section 11, Ch. III Section 8. Markov pure jump processes [4], [1]. **(12L + 3T)**

Unit 3

Stochastic processes (general Theory): Probability spaces appropriate for stochastic processes, Kolmogorov's extension theorem. (only sketch of the proof), separability, progressive measurability and strong Markov property of stochastic processes [3], Ch. 4 Sections 1 and 2. **(12L + 3T)**

Unit 4

Diffusion processes: Definition, elementary properties, infinitesimal parameters, standard process and Dynkin's theorem, continuity and non-differentiability of diffusion processes. Modeling based on diffusion processes, standard Brownian motion, Ornstein-Ulhenbeck process and other processes, transformation of processes, distribution of hitting times and related problems, scale function and speed density, Kolmogorov's backward differential equations, forward differential equations (without proof), transition density and stationary distribution of a diffusion process. **(12L + 3T)**

Note: Section and chapter numbers are from the book by Karlin, S. and Taylor H.M. (1981). *A Second course in Stochastic Processes.* Academic Press.

Books Recommended

1. Adke, S.R. & Manjunath, S. M. (1984). Finite Markov Processes, Wiley.
2. Ash, R.B. and Gardner, M. F. (1975). *Topics in Stochastic Processes*. Academic Press.
3. Athreya, K.B. and Lahiri, S.N. (2006). *Measure Theory and Probability Theory*. Springer.
4. Feller, W. (1969). *An Introduction to Probability Theory*. (Vol. II)
5. Karlin, S. and Taylor H.M. (1981). *A Second course in Stochastic Processes.* Academic Press.

ST P3 Inference for Stochastic Processes (4 Credits)

Unit 1

Inference in Markov chains, estimation of transition probabilities, testing for order of a Markov chain, estimation of functions of transition probabilities, parametric models and their goodness of fit.

Markov sequences, estimation of parameters based on likelihood and conditional least squares, auto-regressive time series. Models for higher order Markov chains. (Raftery's long memory model). Statement of martingale strong law of large numbers and CLT for martingales, CAN property of the MLEs from a general sequence of dependent random variables, Fisher information. Applications to Markov chains and sequences. **(12L + 3T)**

Unit 2

Inference for Poisson process. Likelihood of Poisson and other pure Jump Markov processes from first principles, CAN property of MLEs, testing for a Poisson process, Inference for non-homogeneous Poisson process.

Inference for parametric pure jump processes, such as birth process, birth-death process, birth-death-immigration processes. **(12L + 3T)**

Unit 3

Diffusion processes and their likelihood, properties of estimators (without proof)

Branching processes: ergodic and non-ergodic processes, inconsistency of MLE/moment estimators, properties of estimators on the non-extinction path, estimation of asymptotic distribution theory. **(12L+ 3T)**

Unit 4

Elements of semi-parametric and non-parametric analysis, theory and applications of optimal estimating functions, estimation of transition and stationary density, intensity function of a counting process. Methods based on estimating functions, panel data, introduction to spatial models. **(12L + 3T)**

Books Recommended

1. Adke, S.R. and Manjunath.S.M. (1984). *An introduction to Finite Markov Processes*, Wiley Eastern.
2. Basawa, I.V. and Prakasa Rao, B.L.S.(1980). *Statistical Inference for Stochastic Processes*, Academic Press.
3. Bhat, B. R. (2000). *Stochastic Models: Analysis and Applications*. New Age International.
4. Billingsley, P. (1962). *Statistical Inference for Markov Chains*, Chicago University Press.
5. Guttorp, P. (1991). *Statistical Inference for Branching Processes*, Wiley.
6. Guttorp, P.(1995). *Stochastic Modelling for Scientific Data*, Springer.
7. Prakasa Rao, B.L.S. and Bhat, B.R.(1996). *Stochastic Processes and Statistical Inference*, New Age International.
8. Rajarshi M.B, (2013). *Inference for Discrete Parameter Stochastic Processes*, Springer India.

ST I1: Optimization Techniques

Unit 1

Linear Programming: Review of simplex algorithm and simplex method, artificial variable technique methods. degeneracy, duality in linear programming, duality theorems, dual simplex method with justification.

Integer linear programming problem: pure and mixed integer programming problem, Gomory's all Integer programming method. Fractional cut method- all integer and mixed integer linear programming problem, branch and bound method, dynamic programming, sensitivity. Bellman's optimality principle.

(12L+3T)

Unit 2

Review of transportation and assignment problems: Balance and degeneracy in transportation problem. Transshipment problem, duality theory of testing optimality of solution in transportation problem and transshipment problem, Hungarian method of assignment, maximization, prohibitions and other variations of assignment problems, duality theory of assignment problems.

(12L+3T)

Unit 3

Nonlinear programming: Karush-Kuhn-Tucker conditions, Quadratic programming, Wolfes, Beales and Fletchers algorithms for solving quadratic programming problems. Convex problems, mixed integer models.

(12L+3T)

Unit 4

Networking models: Network flows, maximal flow in the network. Transportation problems, transshipment problems and assignment problems as networking problems. Network scheduling by PERT/CPM Techniques. Resource Analysis in network scheduling. **(12L+3T)**

Books Recommended

1. Bertsekas, D. (1999). *Nonlinear Programming*, 2nd Edn. Athena Scientific.
2. Chong, E. K. P. and Zak, S. (2004). *An Introduction to Optimization*, Wiley.
3. Fletcher, R. (2000). *Practical Methods of Optimization*, Wiley
4. Hadley, G. (1987). *Linear Programming*. Addison-Wesley.
5. Kambo, N.S. (1991). *Mathematical Programming Techniques*. Affiliated East-West press.
6. Panneerselvam, R. (2012). *Operations Research*, 2nd Edn. Prentice Hall of India.
7. Taha, H.A. (1992). *Operations Research*, 5th ed. Macmillan.

ST I2: Statistical Methods for Quality Control

Unit 1

Review: Quality, dimensions of quality, seven SPC tools. Concepts of stable industrial processes, Systematic variation, random variation. Variable & attribute

control charts, $\overline{X}$ chart and R chart. X-MR chart, C chart, U chart, nP chart, P chart. Demerit control chart. CUSUM chart for process mean, CUSUM chart for process variability, Tabular CUSUM. EWMA chart for process mean. EWMA chart for process variability. Comparison of Shewhart control charts with CUSUM chart and EWMA chart.

(12L+3T)

Unit 2

General ideas on economic designing of control charts. Duncan's model for the economic control chart. Concepts of Conforming Run Length (CRL), CRL chart. Properties of CRL chart, Average Run Length (ARL), Average Time to Signal (ATS), ARL and ATS models to obtain the design parameters. Process capability & performance indices, C_p, C_{pk}. Estimation & confidence intervals for estimators of C_p. Connection between proportion of defectives & C_p. Non normal situations.

(12L+3T)

Unit 3

Synthetic control chart to detect increases in fraction non-conforming: Motivation, derivation of ATS. Operation and performance of the synthetic control chart. Synthetic control chart to detect shifts in the process mean: Motivation, derivation of ATS. Operation and performance of the synthetic control chart

Side Sensitive synthetic(SSS) control chart for variables: Motivation. Operation TPM and performance of SSS chart..

'Group Runs' (GR) control chart to detect increases in fraction non-conforming: Motivation, derivation of ATS. Operation and performance of the Group Runs control chart. Comparisons of various control charts. 'Group Runs' GR control chart to detect shifts in the process mean: Motivation, derivation of ATS. Operation and performance of the GR control chart. Comparisons of various control charts. **(12L+3T)**

Unit 4

Multi-Attribute control charts: 'Jolayemi's Multi-Attribute Control Chart' (J-MACC), drawbacks of J-MACC, 'Exact Independent Attribute Control Chart' (E-IACC).

Multivariate control charts for mean vector: The need. Hotelling T^2 chart, Multivariate synthetic control chart.

Acceptance Sampling plans: Single, double & multiple sampling plans for attributes. Curtailed double sampling plans. Operating characteristic functions & other properties of the sampling plan. Use of sampling plans for rectification. Dodge-Romig acceptance sampling plans. Chain sampling plans, Continuous sampling plans CSP-I & CSP – II. Hamaker's conjecture.

Acceptance sampling plan for variables: Designing variable acceptance sampling plans. AQL based sampling plans. Bayesian sampling plans, Minimax regret and other sampling plans. **(12L+3T)**

Books Recommended

1. Guenther, W. C. (1977). *Sampling Inspection in Statistical Quality Control*, Alan Stuart.

2. Levenson, W. (2011). *Statistical Process Control for Real-World Applications*. CRC Press.
3. Montgomery, D. C. (2005). *Introduction to Statistical Quality Control*. Wiley.

ST I3: Statistical Methods for Reliability

Unit 1

Coherent structures, representation of coherent systems in terms of paths and cuts, modules of coherent systems. Reliability of system of independent components, association of random variables, bounds on system reliability, improved bounds on system reliability using modular decompositions.

(12L+1Lab+2T)

Unit 2

Shape of the system reliability function, applications to relay circuits and safety monitoring systems. Notion of aging, life distributions of coherent systems, Distributions with increasing failure rate average arising from shock models, preservation of life distribution classes under reliability operations. Reliability bounds, Mean life series and parallel systems. **(12L+1Lab+2T)**

Unit 3

Classes of life distributions applicable in replacement models, NBU, NBUE, NWU & NWUE classes of life distributions and their implications. Shock models leading to NBU. Age replacement and block replacement policies. Renewal theory useful in replacement models. **(12L+1Lab+2T)**

Unit 4

Replacement policy comparisons, preservation of life distribution classes under reliability operations. Reversed hazard rate, cumulative reversed hazard function, relation between hazard function and reversed hazard function.

(12L+1Lab+2T)

Books Recommended

1. Barlow, R. E. and Proschan F. (1975). *Statistical theory of Reliability and Life testing: Probability Models*. Holt, Rinehart and Winston Inc.
2. Barlow, R. E. and Proschan F. (1996*). Mathematical Theory of Reliability*. John Wiley.
3. Tobias, P. A. and Trindane, D. C. (1995). *Applied Reliability*. Second edition. CRC Press.

ST I4/ST F2: Time Series Analysis

Unit 1

Exploratory time series analysis, tests for trend and seasonality. Exponential and Moving average smoothing. Holt -Winters smoothing. Forecasting based on

smoothing, adaptive smoothing. Time - series as a discrete parameter stochastic process. Auto covariance and autocorrelation functions and their properties, Portmanteau tests for noise sequences, transformation to obtain Gaussian series. **(12L + 3Lab)**

Unit 2

Stationary processes: General linear processes, moving average (MA), auto regressive (AR), and autoregressive moving average (ARMA). Stationarity and inevitability conditions. Nonstationary and seasonal time series models: Auto regressive integrated moving average (ARIMA) models, Seasonal ARIMA (SARIMA) models, Transfer function models (Time series regression)

(12L + 3Lab)

Unit 3

Forecasting in time series models, Durbin-Levinson algorithm, innovation algorithm (without proof). Estimation of mean, auto covariance and autocorrelation functions, Yule-Walker estimation, Estimation of ARIMA model parameters, maximum likelihood method, large sample theory (without proofs). Choice of AR and MA periods, FPE, AIC, BIC, residual analysis and diagnostic checking. Unit-root non stationarity, unit-root tests. **(12L + 3Lab)**

Unit 4

Multivariate Time series model, VAR models, Vector ARMA models. Conditional heteroschedastic models, ARCH and GARCH, properties, examples, estimation & forecasting, extensions of ARCH & GARCH. **(12L + 3Lab)**

Books Recommended:

1. Brockwell, P.J. and Davis, R. A. (2003). *Introduction to Time Series Analysis*, Springer
2. Chatfield, C. (2001). *Time Series Forecasting*, Chapmann & hall, London
3. Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd Ed. Wiley.
4. Hamilton N. Y. (1994). *Time Series Analysis*. Princeton University press.
5. Kendall, M. and Ord, J. K. (1990). *Time Series, 3rd Edn.* Edward Arnold.
6. Lutkepohl, H. and Kratzing, M. (Ed.) (2004). *Applied Time Series Econometrics*, Cambridge University Press.
7. Shumway, R. H.and Stoffer D. S. (2010). *Time Series Analysis & Its Applications*, Springer.
8. Tsay, R. S. (2010). *Analysis of Financial Time Series*, Wiley.

ST- F1: Actuarial Statistics

Unit 1

Future life time random variable, its distribution function and density function, concept of force of mortality, curtate future life time random variable its probability mass function, deferred probabilities, all these functions in terms of international actuarial notation. Analytical laws of mortality such as Gompertz'

law and Makeham's law, Single decrement life table, select and ultimate life table. **(12L + 3T)**

Unit 2

Concept of compound interest rate, discount factor, present value of the money, nominal rate of interest, force of interest. Assurance contracts with level and varying benefits, such as whole life insurance, term insurance endowment insurance. Means and variances of the present value random variables of the payments under these contracts under the assumption of constant force of interest, when the benefit payments are made at the end of year of eath(discrete set up) or when it is paid at the epoch of death(continuous set up). Actuarial present value of the benefit. Net single premiums **(12L + 3T)**

Unit 3

Annuity contracts, annuity certain, discrete annuity, m-thly annuity, continuous annuity, deferred annuity, present values and accumulated values of these annuities.

Continuous life annuity, discrete life annuity, such as whole life annuity, temporary life annuity, n-year certain and life annuity, life annuities with mthly payments.

Present value random variables for these annuity payments, their means and variances. Actuarial present value of the annuity. **(12L + 3T)**

Unit 4

Loss at issue random variable, various principles to decide net premiums for insurance products and annuity schemes defined in unit II and III, fully continuous premiums and fully discrete premiums, True m-thly payment premiums. Extended equivalence principle to decide gross premiums.

Concept of reserve, prospective & retrospective approach. Fully continuous reserve. Fully discrete reserve. **(12L + 3T)**

Books Recommended

1. Bowers, JR. N.L., Gerber, H.U., Hickman, J.C., Jones, D.A. and Nesbitt, C.J. (1997). *Actuarial Mathematics*, 2nd Edn., The Society of Actuaries.
2. Deshmukh S.R. (2009). *Actuarial Statistics: An Introduction Using R*, Universities Press.
3. Harriett, E.J. and Dani, L. L.(1999). *Principles of Insurance: Life,Health, and Annuities*, 2nd Edn., Life Office Management Association.
4. Neill, Alistair (1977). *Life Contingencies*, The Institute of Actuaries.
5. Palande, P. S., Shah, R. S. and Lunawat, M. L. (2003). *Insurance in India - Changing Policies and Emerging Opportunities*, Response Books.

ST F3: Stochastic Models in Finance

Unit 1

Derivatives hedging: forward and future contracts. Markets, prices, arbitrage and hedging

Complete market, market risk and credit risks in the use of derivatives.

Options markets, properties of stock option prices. American and European options. Binomial model: One-step and two-step models, Binomial trees. Risk neutral valuation. **(12L + 3T)**

Unit 2

Behaviour of stock prices: Conditional expectation, martingales, Brownian motion and Geometric Brownian motion, Markov property, Ito integral, Ito/diffusion and mean reverting processes process, Ito Lemma. **(12L + 3T)**

Unit 3

Black Scholes model: Distribution of returns, volatility, risk neutral pricing, equivalent martingale measure, Black-Scholes-Merton differential equation. Estimating volatility (historical data, implied volatility). Options on stock indices, currencies and futures. **(12L + 3T)**

Unit 4

Some exotic equity and foreign exchange derivatives. Greek Letters and hedging. Value-at-risk as a measure of risk. Interest rate derivatives, Black model. Models of the term structure of interest rates: one factor diffusion model, Vesicle, Cox-Ingersoll-Ross and Hull white models **(12L + 3T)**

Books Recommended

1. Baxter, M. and Rennie,.A. (1996). *Financial Calculus*, Cambridge University Press.
2. Bingham, N. and Keisel, R. (1998). *Risk-Neutral Valuation*, Springer.
3. Hull John , (2008). *Options, futures and other derivatives*, International 7th Edn, Pearson Prentice Hall.
4. Ross.S. (2003). *Introduction to Mathematical Finance*, Cambridge University Press.

Additional Reference Books

1. Bodie Z., Kane A., Marcus A. And Mohanty P. (2009). *Investments,* 8th Edn., McGraw Hill.
2. David, R. (2004). *Statistics and Finance: An Introduction*. Springer.
3. Shreve, S. E. (2004). *Stochastic Calculus for Finance I*. Springer.
4. Shreve, S. E. (2004). *Stochastic Calculus for Finance II*. Springer.
5. www.nseindia.com

ST F4/ST B3: Survival Analysis

Unit 1

Concepts of time, order and random censoring. Life distributions - exponential gamma, Lognormal, pareto, linear failure rate. Parametric inference, point estimation, confidence Intervals, scores, tests based on LR, MLE. **(12L+1Lab+2T)**

Unit 2

Life tables, failure rate, mean residual life and their elementary properties. Ageing classes - IFR, IFRA, NBU, NBUE, HNBUE and their duals, Bathtub failure rate. Estimation of survival function - Actuarial estimator, Kaplan - Meier estimator. Estimation under the assumption of IFR/DFR.

(12L+1Lab+2T)

Unit 3

Semi-parametric regression for failure rate - Cox's proportional hazards model with one and several covariates. Competing risk models. Repair models. Probabilistic models. Joint distribution of failure times. Unconditional tests for the time truncated case. **(12L+1Lab+2T)**

Unit 4

Concept of frailty. Shared frailty models. Identifiability of frailty models. Various frailty models. Gamma, positive stable, inverse Gaussian, power variance function, compound Poisson and compound negative binomial shared frailty models. Frailty regression models. Bivariate and correlated frailty models. Additive frailty models.

(12L+1Lab+2T)

Books Recommended

1. Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data*, Chapman and Hall.
3 Deshpande, J.V. and Purohit S.G. (2005). *Life Time Data: Statistical Models and Methods*, Word Scientific.
4 Duchateau, L. and Johnson, P. (2008). *The Frailty Model*. Springer: New York.
5 Gross A.J. and Clark, V. A. (1975) *Survival Distributions: Reliability Applications in the Biomedical Sciences*, John Wiley and Sons.
6 Hanagal, D. D. (2011). *Modeling Survival Data Using Frailty Models*. CRC Press: New York.
7 Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer: New York.
9 Wienke, A. (2011). *Frailty Models in Survival Analysis*, CRC Press: New York.

ST-F5: Multiple Decrement Models in Insurance

Unit 1

Multiple life contracts, joint life status, last survivor status, future life time of the status, curt ate future life time of the status, actuarial present value of the benefit payments, annuity payments and net premiums in a group life insurance contracts. **(12L + 3T)**

Unit 2

Multiple decrement models, distribution of time to decrement and cause of decrement random variables. Multiple decrement table, associated single decrement model, their application in the calculation of monetary functions when cause of death/decrement is involved. Premiums and reserves for insurance contracts involving cause of death/decrement.

(12L + 3T)

Unit 3

Application of multiple decrement models in defined benefit pension plan, calculation of actuarial present value of the benefit payments in a pension plan and determination of the contribution to the pension plan. Various methods of pension funding such as accrued benefit cost method for an individual, accrued benefit cost method for a group, aggregate actuarial cost method

(12L + 3T)

Unit 4

Multistate Markov models for cash flows contingent on competing risks, actuarial present value of the cash flows, determination of premium in disability income insurance contract in employee benefit schemes, determination of premium in continuing care retirement communities model in health insurance.
Stochastic interest rate, various models such as random interest scenario, parametric models, time series models and its application in the calculation of monetary functions in a variety of insurance contracts. **(12L + 3T)**

Books Recommended

1. Bowers, JR. N.L., Gerber, H.U., Hickman, J.C., Jones, D.A. and Nesbitt, C.J. (1997). *Actuarial Mathematics*, Second Edition, The Society of Actuaries.
2. Deshmukh S.R. (2012). *Multiple Decrement Models in Insurance: An Introduction Using R*, Springer India.
3. Harriett, E.J. and Dani, L.L.(1999).*Principles of Insurance: Life, Health, and Annuities*, Second Edition, Life Office Management Association
4. Neill, Alistair (1977). *Life Contingencies*, Institute of Actuaries, London.

ST B2: Statistical Analysis of Clinical Trials

Unit 1

Introduction to clinical trials: need and ethics of clinical trials, bias and random error in clinical studies, conduct of clinical trials, overview of Phase I-IV trials, multi-center trials. Data management: data definitions, case report forms, database design, data collection systems for good clinical practice. Bioavailibility, pharmacokinetics and pharmacodynamics, two-compartment model.

(12L+1Lab+2T)

Unit 2

Design of clinical trials: parallel vs. cross-over designs, cross-sectional vs. longitudinal designs, objectives and endpoints of clinical trials, design of Phase I trials, design of single-stage and multi-stage Phase II trials. Design and monitoring of Phase III trials with sequential stopping, design of bio-equivalence trials. Inference for 2x2 crossover design: Classical methods of interval hypothesis testing for bioequivalence, Bayesian methods, nonparametric methods. **(12L+1Lab+2T)**

Unit 3

Power and sample size determination, multiplicative (or log-transformed) model, ML method of estimation, assessment of inter and intra subject variabilities, detection of outlying subjects. Optimal crossover designs: Balaam's design, Two-sequence dual design. Optimal four period designs. Assessment of bioequivalence for more than two drugs, Williams design. **(12L+1Lab+2T)**

Unit 4

Designs based on clinical endpoints: Weighted least squares method, log-linear models, generalized estimating equations. Drug interaction study, dose proportionality study, steady state analysis. Interim analysis and group sequential tests, alpha spending functions. Analysis of categorical data.

(12L+1Lab+2T)

Books Recommended

1. Chow S.C. and Liu J.P.(2009). *Design and Analysis of Bioavailability and bioequivalence*. 3rd Edn. CRC Press.
2. Chow S.C. and Liu J.P. (2004). *Design and Analysis of Clinical Trials*. 2nd Edn. Marcel Dekkar.
3. Fleiss J. L.(1989). *The Design and Analysis of Clinical Experiments*. Wiley.
4. Friedman L. M.Furburg C. Demets D. L.(1998). *Fundamentals of Clinical Trials*, Springer.
5. Jennison .C. and Turnbull B. W. (1999). *Group Sequential Methods with Applications to Clinical Trails*, CRC Press.
6. Marubeni .E. and Valsecchi M. G. (1994). *Analyzing Survival Data from Clinical Trials and Observational Studies*, Wiley.

ST B4: Statistical Methods in Microarray Data Analysis

Unit 1

Background of Microarrays and Normalization techniques

Introduction to Biology relevant to microarray experiment. Microarray experimental set up and quantification of information available from microarray experiments.

Data cleaning, transformation of data.

Between array & within array normalization, in particular quantile and LOWESS normalization, stage wise normalization.

Concordance coefficient and its role in normalization.

(12L + 3T)

Unit 2

Statistical Inference procedures in comparative experiments

Inference procedures for single channel microarray data. Application of two sample t –test. Tests for validating assumptions of two sample t-test. Application of Welch test and Wilcoxon rank sum test.

Inference procedures for two channel microarray data. Application of paired t – test. Tests for validating assumptions of paired t-test. Application of Wilcoxon signed rank test.

Inference procedures for comparing more than two types of mRNA samples in single channel or two channel microarray experiments. Application of one way ANOVA F test, one way ANOVA Welch F test, Kruskal-Wallis test, pairwise t-test, pairwise Welch test and pairwise Wilcoxon rank sum test. Strip charts and its role to decide the profile of differentially expressed genes

(12L + 3T)

Unit 3

Multiple hypotheses testing problem and Principal component analysis

Multiple hypotheses testing problem. Adjustments for multiple hypotheses testing, adjusted p-values.

False discovery rate and its application to microarray data analysis.

Principal component analysis for microarray data, scree plot, plot of scores to display sets of differentially expressed genes. Singular value decomposition of a rectangular matrix and the concept of ballot. Its application to microarray data analysis.

(12L + 3T)

Unit 4

Cluster analysis and Logistic regression

Hierarchical cluster analysis of microarray data to identify groups of genes and outlying genes

K - means cluster analysis of microarray data to identify groups of genes

Application of logistic regression for microarray data. Concept of AIC and BIC and its role to identify marker genes.

(12L + 3T)

R software will be heavily used in applications of all the statistical methods to microarray data to identify differentially expressed genes in two or more biological samples.

Books Recommended

1. Amartunga D. and Cabrera J. (2004). *Exploration and Analysis of DNA Microarray and Protein Array Data*. Wiley.
2. Deshmukh S.R. (2007). *Microarray Data: Statistical Analysis Using R*, Narosa.
3. Draghici, S. (2003). *Data Analysis Tools for DNA Microarrays*, Chapman and Hall/CRC.
4. Dov, S. (2003). *Microarray Bioinformatics*, Cambridge University Press,
5. McLachlan, G.J.; Do, K.A. and Ambroise, C. (2004). *Analyzing Microarray Gene Expression Data*, Wiley.
6. Simon, R.M. ; Korn, E.L. ; McShane, L.M. ; Radmacher, M.D. ; Wright, G.W. and Zhao, y. (2003). *Design and Analysis of DNA Microarray Investigations*. Springer.
7. Speed, T. (2003). *Statistical Analysis of Gene Expression Microarray Data*, Chapman and Hall/CRC.

ST-C2: Computer Intensive Statistical Methods

Note: It is recommended that this course be conducted in Computer Laboratory. R will be used for computing purpose.

Unit 1

Resampling Techniques: Re sampling paradigms, bias-variance trade-off. Bootstrap methods, estimation of sampling distribution, confidence interval, variance stabilizing transformation. Jackknife and cross-validation. Jackknife in sample surveys. Jackknife in regression under heteroscedasticity. Permutation tests. **(12L + 3T)**

Unit 2

Missing Values and Imputations Techniques: Missing values and types of missingness, imputations methods for missing values, single and multiple imputations.

EM Algorithm and Applications: EM algorithm for incomplete data, EM algorithm for mixture models, EM algorithm for missing values, stochastic EM algorithm.

(12L+3T)

Unit 3

Smoothing techniques: Kernel estimators, nearest neighbor estimators, orthogonal and local polynomial estimators, wavelet estimators. Splines. Choice of bandwidth and other smoothing parameters.

(12L+3T)

Unit 4

Bayesian computing, Markov Chain Monte Carlo. Simulation using MCMC, Particle filtering, MCMC methods for missing values. **(12L+3T)**

Books Recommended

1. Buuren, Stef van (2012). *Flexible Imputation of Missing Data.* Chapman and Hall.
2. Chihara, L. and Hesterberg, T. (2011) *Mathematical Statistics with Resampling and R*. Wiley.
3. Davison, A.C. and Hinkley, D.V. (1997) *Bootstrap methods and their Applications*. Chapman and Hall.
4. Efron, B. and Tibshirani. R.J. (1994); *An Introduction to the Bootstrap*. Chapman and Hall.
5. Christensen R, Johnson, W., Branscum A. and Fishman, G.S. (1996) *Monte Carlo: Concepts, Algorithms, and Applications*. Springer.
6. Gilks, W. R., Richardson, S., and Spiegelhalter, D. (eds.) (1995) *Markov Chain Monte Carlo in Practice*. Chapman and Hall.
7. Good, P. I. (2005) *Resampling Methods: A Practical Guide to Data Analysis*. Birkhauser Bosel.
8. Hanson T. E. (2011). *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*, Chapman & Hall.
9. Jim, A. (2009). *Bayesian Computation with R*, 2nd Edn, Springer.
10. Kennedy W. J. & Gentle J. E. (1980) *Statistical computing*. Marcel Dekker.
11. McLachlan, G.J. and Krishnan, T. (2008) *The EM Algorithms and Extensions*. Wiley.
12. Rubinstein, R.Y. (1981); *Simulation and the Monte Carlo Method*. Wiley.
13. Shao J. and Tu, D. (1995); *The Jackknife and the Bootstrap*. Springer Verlag.
14. Tanner, M.A. (1996); *Tools for Statistical Inference*, Third edition. Springer.

ST-E9: Financial Econometrics

Unit 1

Unit roots, Cointegration and VAR models: difference stationary and trend stationary processes. Testing for unit roots: the DF, ADF, PP and KPSS test statistics, VAR, ML estimation Granger causality, Cointegrating VAR, Applications to the PPP (purchasing power parity), Applications to the net present value model of stock prices, market microstructure and the efficient market hypothesis, examples & data analysis **(12L+3T)**

Unit 2

Frequency domain analysis of time series: Fourier transform and discrete Fourier Transform (DFT), periodicity, Spectral density, periodogram and DFT, spectral representation, inference, examples & data analysis **(12L+3T)**

Unit 3

Structural equation modeling, state-space models, Kalman filter: State-space representations, the basic structural model, state-space representation of ARIMA models, The Kalman recursions, estimation for State-Space models, Generalized state-space models, parameter & observation-driven models, examples & data analysis. **(12L+3T)**

Unit 4

Stochastic volatility models: Volatility definition and estimation, Volatility forecast evaluation, Stochastic volatility models, MCMC approach, examples.
Modeling regime shifts - Markov-switching models: Representation (AR & MA), estimation, diagnostic checking, forecasting, examples & data analysis

(12L+3T)

Recommended Books:

1. Ait-Sahalia, Y. & Hansen, L. P. (Ed.) (2010). *Handbook of Financial Econometrics: Tools and Techniques*, Vol.1 & Vol. 2, Elsevier.
2. Campbell J., Lo A. & McKinley C. (1997). *The Econometrics of Financial Markets*. Princeton University press.
3. Hamilton (1994). *Time Series Analysis*. Princeton University press.
4. Knight, J. and Satchell, S. (2007). *Forecasting Volatility in the Financial Markets,* 3rd Edn., Elsevier.
5. Lutkepohl, H. and Kratzing, M. (Ed.) (2004). *Applied Time Series Econometrics*, Cambridge University Press.
6. Poon Ser-Huang (2005). *A Practical Guide to Forecasting Financial Market Volatility*, Wiley.
7. Rachev. S. T., Mittnik, S., Fabozzi, F. J. , Focardi, S. M. and Jasic, T. (2007). *Financial Econometrics:From Basics to Advanced Modeling Techniques,* Wiley.
8. Ruppert, D. (2004). *Statistics and Finance: An Introduction*, Springer.
9. Shephard, N. (2004). *Stochastic Volatility: Selected Readings*, Oxford University Press.
10. Soderlind, P. (2010). *Lecture Notes in Financial Econometrics*, University of St. Gallen.
11. Tsay, R. S. (2010). *Analysis of Financial Time Series*, 3rd Edn. Wiley.
12. Wang, P. (2003). *Financial Econometrics: Methods and Models*, Routledge.

ST C4: Statistical Learning and Data Mining

Note: It is recommended that this course be conducted in Computer Laboratory.

Unit 1

Supervised Learning: Linear methods for classification, linear discriminant analysis (LDA), logistic regression, Bayes classifier, nearest neighbor classifier. Packages in R for these methods. **(12L + 3T)**

Unit 2
Neural network (NN), support vector machine (SVM). Packages in R for these methods. **(12L + 3T)**

Unit 3
Regression and classification trees (CART). Assessment and model selection: Bias-variance trade off, training error rate, AIC, BIC, CIC, DIC (information criterion), cross-validation. Ada boosting. **(12L + 3T)**

Unit 4
Unsupervised learning: Clustering procedures- k-means, hierarchical, self-organizing map, EM algorithm. Feature selection: principal component analysis. Association rules. Software packages for these methods. **(12L + 3T)**

Books Recommended

1. Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth and Brooks.
2. Daniel T. Larose, (2006). *Data Mining Methods and Models*. Wiley.
3. Galit Shmueli, Nitin Patel, Peter Bruce, (2010). *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XL Miner* , Wiley
4. Hastie T., Tibshirani R. and Friedman J. H., (2003). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer.
5. Mitchell Tom, (1997). Machine Learning. McGraw-Hill.
6. Ripley,B. D. (1996). *Pattern Recognition and Neural Networks.* Cambridge University Press.

ST B6/ST C6 : Statistical Methods for Bio-Computing

Unit 1
Type of genetic data:- Molecular and morphological data. Differences and advantages of molecular data on, morphological data, Character data and distance data, their relative merits and demerits.
Concept of entropy, entropy as a measure of uncertainty, entropy of single and combined scheme/s, Measure of information content based on entropy. Relative entropy its similarity with likelihood ratio. Applications of these to biological sequences.

(12L + 3 L T)

Unit 2 (Alignment of biological sequences): Pairwise and local alignment of biological Sequences (DNA/protein sequences). How biological sequences are different from mathematical sequences? The scoring matrices for alignment algorithms PAM and BLOSUM matrices. Algorithm for global alignment (Needleman–Wunch algorithm). Local alignment algorithms (Smith - Waterman) Gap Model, dynamic programming algorithms for alignment with gaps such as

linear gap model, affine gap model. Introduction to heuristic alignment algorithms such as BLAST, FASTA.

(12L + 3 T/L)

Unit 3

Molecular phylogeny Analysis: Tree of life, gene and species tree. Distance based methods for reconstruction of phylogenetic tree such as UPGMA, weighted UPGMA, transformed distance method, nearest – neighbor joining method. Comparison of trees generated using different distance function Requisites of a good distance function.

Character based methods for molecular phylogeny, maximum likelihood method and maximum parsimony method. Assessing trees via bootstrap. Probabilistic approach to phylogeny. Probabilistic models of evolution, Felsentein's algorithm for likelihood computation. Juke – Canter model and Kimura and other probabilistic models for evolution.

(12L+4T)

Unit 4

Applications of Markov and Hidden Markov models to biological sequence Analysis.

Markov chain as a classifier, use of Markov chain Model for demarcation of a region in Biological sequence analysis. Application of these in genetic sequence analysis such as detection of CPG Island. Testing whether given stretch of sequence is coming from CPG Island (use of Markov model for discrimination) Markov model based classification & clusterization, testing order of a Markov model, testing homogeneity of two Markov models, Use of these test to design clustering algorithm. Hidden Markov/chains. Difference between these and simple Markov chains. Analysis of Hidden Markov Models/chains. Verterbi's algorithm, Forward and backward algorithm for hidden Markov model. Parameter estimation in hidden Markov model when path is known as well as unknown, Baum – Welch algorithm.

(12L + 3T)

Recommended Books

1. Alexander Isaac: (2001). *Introduction to Mathematical Methods Bioinformatics*. Springer.
2. Durbin R., Eddy S. Krogh A. Michelson G. (1998). *Biological Sequence Analysis*, Cambridge University Press.
3. Robin S., Rudolph F, Schboth S. (2003) *DNA Words and models Statistics of Exceptional Words*, Cambridge University Press.