MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL
Syllabus of M. Sc. Applied Statistics and Analytics (In-House Programme)
Duration: 2 Years; **Level**: Post graduation; **Type**: Degree
(Effective from academic session 2020-21)

Semester-2

Regression for Predictive Model Building	**Paper Code: MSASA201**
	4 Credits, 100 marks

Course Objectives:

1. To understand notion of statistical model building
2. To understand different aspects of regression diagnostics and their remedies
3. To use different kinds of statistical models for suitable data
4. To learn modern era regression based on decision tree and random forest

Module 1 (20L)

Building a regression model: Transformations of data, Stepwise regression, Model selection (adjusted R2, cross validation and Cp criteria, AIC, BIC). Multicollinearity – detection and remedial measures. Dummy variables, piecewise regression, splines and scatter plot smoothing.

Detection of outliers and influential observations: residuals and leverages. Checking for normality:Q-Q plots, Normal Probability plot, Shapiro-Wilks's test. Heteroscedasticity and Autocorrelation – detection and remedies. Bayesian Regression, Robustification of regression model.

Module 2 (8L)

Generalized Linear Models: Introduction, Components of a GLM, Maximum Likelihood estimation, Deviance, Introduction to categorical data, Contingency tables and discrete data analysis, Odds ratio in 2x2 table, Partial and conditional odds ratio in multiway tables.

Module 3 (8L)

Different regression models (Logistic regression, Count regression, Polytomous regression), context, goodness of fit tests to real data.

Module 4 (8L)

Decision tree, algorithm for tree-based models, random forest, variable interest, interaction among predictors

References

1. Applied Regression: An Introduction, Colin Lewis-Beck and Michael Lewis-Beck, Second Edition, 2015, SAGE
2. Regression Modelling Strategies, Frank E. Harrell, Jr., Second Education, Springer, 2015
3. David A. Belsley, Edwin Kuh, Roy E. Welsch, Regression Diagnostics – identifying Influential data & sources of collinearity, 2013
4. Johnston J, Econometric Methods, Edition 4, McGraw-Hill

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL
Syllabus of M. Sc. Applied Statistics and Analytics (In-House Programme)
Duration: 2 Years; **Level**: Post graduation; **Type**: Degree
(Effective from academic session 2020-21)

Optimization Techniques and Soft Computing	**Paper Code: MSASA202**
	4 Credits, 100 marks

Course Objectives:

1. To transform real life minima/maxima problems into optimization framework.
2. To learn efficient computational procedures to solve linear and non-linear optimization problems.
3. To acquire an idea about the various direct and indirect search methods.
4. To learn applicability of different queuing models in real life cases.
5. To learn the different soft computing techniques including Fuzzy logic and evolutionary algorithms.

Module-1 **(6L)**

Introduction: Concept of optimization – Classification of optimization problems.
Linear Programming: Examples of linear programming problems – Formulation, Simplex methods including Charne's Penalty Method, Two Phase Method and Degeneracy, Dual simplex method, Sensitivity analysis.
Solution of the Transportation and Assignment problems, Shortest route problem.

Module-2 **(9L)**

Non-Linear Optimization Techniques:
Unconstrained Optimization: Maximization and minimization of convex functions, Necessary and sufficient conditions for local minima, Speed and order of convergence,

Univariate search methods: Direct Search Methods-Fibonacci and Golden Section Search, Descent Methods-Steepest Descent, Fletcher Reeves, Newton's and Conjugate Gradient methods.

Unconstrained Optimization Techniques: Necessary and sufficient condition for Kuhn – Tucker conditions for optimality, Lagrange's Multiplier Method, Gradient projection method, Penalty function methods.

Module-3 **(5L)**

Queueing Theory: Introduction of the queuing system, Various components of a queueing system. Pure Birth Process; Pure Death Process, Birth and Death Process, M/M/1 and M/M/c models, M/M/c model with finite waiting space, and models with a finite source of customers (machine interference problem). Semi-Markovian queueing systems: M/G/1, M/G/1 with service vacations and G/M/1.

Module-4 **(6L)**

Introduction to Soft Computing: Concept of computing systems, "Soft" computing versus "Hard" computing, Characteristics of Soft computing, Some applications of Soft computing techniques.

Fuzzy logic**:** Introduction to Fuzzy logic, Fuzzy sets and membership functions, Operations on Fuzzy sets, Fuzzy relations, rules, propositions, implications and inferences, Some applications of Fuzzy logic, Fuzzy Inference Systems,

Module-5 **(8L)**

Genetic Algorithm (GA): Differences and similarities between conventional and evolutionary algorithms, working principle, Genetic Operators- reproduction, crossover, mutation, Solving single-objective optimization problems using GAs.

Artificial Neural Networks: Biological neurons and its working, Simulation of biolgical neurons to problem solving, Different ANNs architectures, Applications of ANNs to solve some real life problems.

Module-6 **(6L)**

Optimization in Finance: Short term financing, asset pricing and arbitrage, portfolio selection and asset allocation, the fundamental theorem of asset pricing, arbitrage detection using LP.

Supply chain optimization problems:
Introduction to Supply-Chain Optimization, Sequencing and scheduling problems in production planning, Classical machine scheduling problems.

Formulation of n -period supply-chain problem under certainty. Minimization of n period costs subject to stock conservation constraints.

Myopic Supply Policy with Stochastic Demands.

References

1. H. A. Taha. “operations Research, An Introduction”, PHI, India, 2002.
2. S. S. Rao, “Engineering Optimization: Theory and Practice”, 4th Edition, John Wiley & Sons (2009). • Kwang Y. Lee,
3. D. Gross and C. M. Harris, Fundamentals of Queueing Theory, Wiley, 2018
4. Fuzzy Logic with Engineering Applications (3rd Edn.), Timothy J. Ross, Willey, 2010.
5. Foundations of Neural Networks, Fuzzy Systems, and Knowldge Engineering, Nikola K. Kasabov, MIT Press, 1998.
6. An Introduction to Genetic Algorithms, Melanie Mitchell, MIT Press, 2000.
7. Genetic Algorithms in Search, Optimization and Machine Learning, David E. Goldberg, Pearson Education, 2002.
8. Soft Computing, D. K. Pratihar, Narosa, 2008.
9. Neural Networks and Learning Machines, (3rd Edn.), Simon Haykin, PHI Learning, 2011.
10. Optimization Methods in Finance, Gerard Cornuejols and Reha Tutuncu,Carnegie Mellon University, Pittsburgh, PA 15213 USA, January 2006
11. Lectures in Supply-Chain Optimization Arthur F. Veinott, Jr., Department of Management Science and Engineering, Stanford University, Stanford, California 94305

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL
Syllabus of M. Sc. Applied Statistics and Analytics (In-House Programme)
Duration: 2 Years; **Level**: Post graduation; **Type**: Degree
(Effective from academic session 2020-21)

Stochastic Processes and its Application	**Paper Code: MSASA203**
	4 Credits, 100 marks

Course Objectives:

1. To gain the knowledge of probabilistic methods in engineering and scientific application.
2. To learn applicability of essential mathematical tools for handling random processes.
3. To familiarize the learner with the stochastic simulation techniques.
4. To acquire the ability for applications of probabilistic and stochastic methods in modern engineering problems.

Module-1 **(10L)**

Branching Processes: Brief overview of modelling -- deterministic/stochastic; discrete time / continuous time. review of discrete time branching process, extinction probabilities and asymptotic behaviour, Brief excursion to continuous time branching process, general two-type birth-death branching process, Branching process with general lifetime variable (Bellman-Harris process).

Module-2 **(10L)**

Modelling in Genetics: Brief review of genetics, including the Hardy-Wienberg laws, their ramifications including mutation and fitness coefficient, Inbreeding and changes of coefficient of inbreeding over generations, Markovian models: Sib mating, Wright-Fisher, Moran, Kimura models, Wright-Fisher model with varying generation sizes, Hidden Markov models.

Module-3 **(12L)**

Random Processes: Introduction and classification of Random Processes, Stationarity and Independence of random processes, First and higher order Stationary Processes.

Markov Processes: Discrete Markov chains and Markov processes, the Markovian property, Chapman-Kolmogorov's theorem and the classification of Markov processes. Transition probability, Transition intensity, Forward and backward equations, Stationary and asymptotic distributions, Convergence of Markov chains, birth-and-death processes, absorption probabilities, time to absorption, renewal theory, martingales, Brownian motion and diffusion processes.

Module-4 **(8L)**

Epidemic Modelling: Simple and general epidemics - both deterministic as well as stochastic. Threshold theorems (without proof). Greenwood, Reed-Frost models, Neyman-Scott models of spatial spread of epidemics.

References:

1. S. Karlin, H. M. Taylor: A First Course in Stochastic Processes. Academic Press
2. N. T. J. Bailey: The Elements of Stochastic Processes. Wiley
3. D. R. Cox and H. D. Miller: The Theory of Stochastic Processes. Chapman and Hall/CRC
4. N. T. J. Bailey: The mathematical theory of infectious diseases and its applications, London: Griffin
5. J. Medhi: Stochastic Processes, Wiley Eastern Limited

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL
Syllabus of M. Sc. Applied Statistics and Analytics (In-House Programme)
Duration: 2 Years; **Level**: Post graduation; **Type**: Degree
(Effective from academic session 2020-21)

Time Series Analysis and Forecasting Methods	**Paper Code: MSASA204**
	4 Credits, 100 marks

Course Objectives:

1. To interpret time series data in the context of analytics
2. To apply non stochastic and stochastic aspects of time series through different models
3. To learn application of time series modelling in share market or other arenas where volatility is frequent.
4. To learn a combined analysis both involving time series as well as regression.

Module-1 **(10L)**

Introduction: Classical Models, Smoothing Techniques – exponential and Holt-Winters methods.

Module-2 **(10L)**

Evolutionary and Stationary time series. Autocorrelation and partial autocorrelation functions. AR, MA, ARMA, ARIMA, SARIMA

Module-3 **(14L)**

Box-Jenkins Model. Volatility: ARCH, GARCH models and their variants. Multivariate Time Series Models: VAR and VARMA models. Forecasting Methods, Time series regression models

Module-4 **(6L)**

Analysis in the Frequency Domain: The Spectrum and Periodogram Analysis.

References:

1. C. Chatfield: The Analysis of Time Series – An Introduction, Chapman Hall
2. G.E.P. Box, G.M. Jenkins & G.C.Reinsel: Time Series Analysis – Forecasting & Control, Wiley
3. P.J. Brockwell & R.A. Davis: Introduction to Time Series Analysis and Forecasting, Wiley
4. Ruey S. Tsay: Analysis of Financial Time Series, Wiley

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL
Syllabus of M. Sc. Applied Statistics and Analytics (In-House Programme)
Duration: 2 Years; **Level**: Post graduation; **Type**: Degree
(Effective from academic session 2020-21)

Evolution of Statistical Thinking	**Paper Code: MSASA205**
	Audit Course

Course Objectives and outlines:

We all begin life with naïve realism: the doctrine that things are how they appear. But things are often not what they seem. We like to believe that effects are completely determined by comprehensible causes. But a lot of what happens in life is as much a result of random factors as is due to known ones. Thus, our past is not so easy to understand, nor is our future easy to predict; and in both pursuits we benefit from looking beyond the superficial explanation. This course will survey the history of the struggle of the human mind to comprehend and formulate the randomness and uncertainty inherent in life, and in doing so, underscore the mathematics invented in this process. Like archaeology, it will present the tools needed to identify the footprints of chance.

The contribution of Jakob Bernoulli will be a central point of reference in this journey. He had shown that through mathematical analysis one could learn how the inner hidden probabilities that underlie natural systems are reflected in the data those systems produce. However, Bernoulli left for his posterity to solve the question of how to infer, from the data produced, the underlying probability of events.

The fact that human intuition is ill suited to situations involving uncertainty was known as early as the 1930s. However, that knowledge for the most part have not trickled down from academic circles to the popular psyche. This course is an attempt to remedy that. It aspires to enable and inspire the students to integrate probabilistic and statistical thinking in the conduct of their daily lives, and to not be fooled by randomness.

This course is about the principles that govern 'chance' inherent in life, the development of those principles, and how they play out in politics, business, medicine, economics, sports, leisure, and other indispensable areas of human affairs. It is also about the way we make choices and the processes that lead us to make mistaken judgments and poor decisions when confronted with the unavoidable uncertainty and randomness of life. The course proceeds by taking the mind to the following tourist spots of inquiry:

- The hidden role of chance … when human beings can be outperformed by a rat.
- The basic principles of probability and how they are abused … why a good story is less likely to be true than a flimsy explanation.
- A framework for thinking about random situations … from a gambler in plague-ridden Italy to Let's Make a Deal.
- How to count the number of ways in which events can happen, and why it matters … the mathematical meaning of expectation.
- The extent to which probabilities are reflected in the results we observe … Zeno's paradox, the concept of limits, and beating the casino at roulette.
- How to adjust expectations in light of past events or new knowledge … mistakes in conditional probability from medical screening to the OJ Simpson trial and the prosecutor's fallacy.
- The meaning and lack of meaning in measurements … the bell curve and wine ratings, political polls, grades, and the position of planets.

- How large numbers can wash out the disorder of randomness … or why 200,000,000 drivers form a creature of habit.
- Fooled by the regularities in chance events … can the success of Wall Street gurus be random?
- Risk and uncertainty … and the deep danger of confusing the two.

References

1. The Drunkard's Walk – How Randomness Rules Our Lives" by Leonard Mlodinow, Penguin, 2009.
2. The Art of Statistics – Learning from Data" by David Spiegelhalter, Pelican, 2020.

Regression and Time Series Laboratory	**Paper Code: MSASA291** **4 Credits, 100 marks**

Chapter	Name of the topic	Hours
01	Regression diagnostics	4
02	Remedial measures and model validation	6
03	Non-Stochastic and stochastic analysis of time series data	6
04	Forecasting and regression time series	4

Optimization Techniques and Stochastic Process Laboratory	**Paper Code: MSASA292** **4 Credits, 100 marks**

Chapter	Name of the topic	Hours
01	Optimization technique through software	5
02	Applications on queuing, soft computing and fuzzy logic	5
03	Applications on Markov Chain and hidden Markov models	5
04	Epidemic Modelling and Black Scholes model	5