Admission to the Ph.D. Programme in Statistics: 2014

Procedure:

1. The conditions for eligibility will be guided overall by the rules specified in the notification titled **University of Calcutta (Regulations for the Degree of Doctor of Philosophy, Ph.D.), Regulations 2009** (http://www.caluniv.ac.in/Phd_Dlit/PhD%20rules.pdf)
2. Eligibility: Candidates with an M.Sc. or equivalent degree in Statistics and allied subjects from any UGC recognized University/Institute are eligible to apply for admission in the Ph.D. programme.
3. The admission procedure consists of a written test followed by an interview for candidates successful in the same. Those who have qualified NET / SET (Mathematical Sciences) / GATE (Mathematics) or hold a UGC Teachers' Fellowship or already obtained M.Phil. in Statistics and allied subjects or M. Tech. (QR&OR) degree of ISI prior to the application deadline will be exempted from the written examination but will have to appear in the interview.
4. The written test is common to the Ph.D. and M.Phil. programmes for the year 2014. A candidate who qualifies for the Ph.D. programme is automatically eligible to join the M. Phil. programme instead, if he/she chooses. However, for exercising this choice, such a candidate must have applied also for the M. Phil. programme by the relevant deadline. The two programmes cannot be pursued simultaneously.
5. Number of seats in the Ph.D. programme: 58
6. Reservations will be applicable as per existing rules.

Date of Advertisement	:	
Last date of submission of application form	:	September 26, 2014
Date of common written test	:	October 14, 2014 (12 noon - 2 p.m.)
Result of common written test	:	October 21, 2014
Date of Interview	:	November 05, 2014 (from 12 noon)
Date of publication of selection list	:	November 17, 2014

Please note that candidates who are eligible for waiver of the written test are also required to complete and submit the application form by the above deadline.

Course Work (PhD Programme):

One Semester Course Work of 20 credits as follows:

1. **Literature Review and seminar : 4 credits**
2. **Seminar Presentation : 4 credits**
3. **Research Methodology : 4 credits**
4. **Evolution of Statistics : 4 credits**
5. **Statistical Computing : 4 credits**

Structure of the written examination:

1. There will be 25 multiple choice questions each carrying 2.5 marks out of which one has to answer 20 questions. If a candidate answers more than 20, only the first 20 answered will be evaluated.
2. There will be 15 short answer type questions of 5 marks each out of which one has to answer 10 questions. If a candidate attempts more than 10, only the first 10 attempted will be evaluated.
3. Pass marks for the test will be announced in due course.
4. Candidates successful in the written examination would have to compete with other eligible candidates who have already cleared NET / SET / GATE / M. Phil / M. Tech. (QR&OR) at the interview stage. The list of finally selected candidates would be posted in the University website and Departmental Notice Board.
5. Candidates selected for the final interview will be required to specify his/her areas of interest by a specified date before the interview. However, the selection committee may, at its discretion, require a candidate to opt for a topic/area other than his/her initial choice before admitting him/her into the Ph.D. programme. The final date for submitting the brief will be announced along with the intimation for the interview.

Detailed Syllabus for common M.Phil-PhD Entrance Examination:

Real Analysis

Real number system, cluster points of sets, closed, open and compact sets, Bolzano-Weierstrass property, Heine-Borel property. Sets of real vectors. Sequences and series, convergence. Real valued functions. Limit, continuity and uniform continuity. Differentiability of univariate functions. Mean value theorems. Extrema of functions. Riemann integrals. Improper integrals. Sequences and series of functions, uniform convergence. Power series. term by term differentiation and integration,

Probability

Fields, sigma-fields and generators, semifields, Borel sigma-field on R and R^k. Monotone classes, Measurable functions and properties, compositions; product sigma-fields, Borel sigma-field on Euclidean spaces.

Measures, finite, sigma-finite measures. Probability measures, properties. Independence of events, Borel-Cantelli lemmas.

Extensions of measures, Lebesgue measure on R and R^k.

induced measures. Random variables, Distribution functions, measures in R and R^k. Probability distributions. Discrete and absolutely continuous distributions. probability densities.

Convergence in probability and almost sure.

Integration: simple, nonnegative, general measurable functions, integrability, MCT, DCT, Fatou's lemma. Change of variables. Holder's and Minkowski's inequalities. Expectations, moments. Jensen's inequality. Product measures. Fubini's theorem.

Independence of random variables. Sums, variances, covariances. Second Borel-Cantelli lemma. Kolmogorov's 0-1 law. Weak and strong laws of large numbers. Kolmogorov's inequality.

Convergence in distribution. Integration of complex-valued functions, characteristic functions. Inversion and Continuity theorems. Central Limit Theorems.

Lp-convergence of random variables. Connections between various modes of convergence (in distribution, in probability, L_p, almost sure).

Absolute continuity and singularity of measures. Radon-Nikodym theorem (statement).

Linear Algebra and Linear Programming

Vectors and Matrices: Vector spaces and subspaces, Linear dependence and independence, span, basis, orthogonality and orthonormality,

Matrix algebra

Linear programming: Graphical Solution and Simplex Algorithm

Sampling Distributions

Non-central χ^2, t & F distributions – definitions and properties.Distribution of quadratic forms – Cochran’s theorem.

Large Sample Theory

Scheffe's theorem, Slutsky's theorem. Asymptotic normality, multivariate CLTs, delta method. Glivenko-Cantelli Lemma

Asymptotic distributions of sample moments and functions of moments, Asymptotic distributions of Order Statistics and Quantiles. Consistency and Asymptotic Efficiency of Estimators, Large sample properties of Maximum Likelihood estimators. Asymptotic distributions and properties of Likelihood ratio tests, Rao’s test and Wald’s tests in the simple hypothesis case.

Statistical Inference

Sufficiency & completeness, Notions of minimal sufficiency,bounded completeness and ancillarity, Exponential family.Point estimation : Bhattacharya system of lower bounds to variance of estimators. Minimum variance unbiased estimators – Applications of Rao – Blackwell and Lehmann – Scheffe theorems.Testing of Hypothesis : nonrandomized and randomized tests, critical function, power function. MP tests – Neyman – Pearson Lemma. UMP tests. Monotone Likelihood Ratio families. Generalized Neyman – Pearson Lemma. UMPU tests for one parameter families. Locally best tests. Similar tests.

Neyman structure. UMPU tests for composite hypotheses.
Confidence sets: relation with hypothesis testing. Optimum parametric confidence intervals. Sequential tests. Wald's equation for ASN. SPRT and its properties – fundamental identity. O.C. and ASN. Optimality of SPRT (under usual approximation).

Linear Models

Gauss Markov Model: Estimable function, error function, BLUE, Gauss Markov theorem. Correlated set-up, least squares estimate with restriction on parameters.
Linear Set, General linear hypothesis –related sampling distributions, Multiple comparison techniques due to Scheffe and Tukey.
Analysis of variance: Balanced classification, Fixed Effects Model, Random Effects Model and Mixed Effects Model; Inference on Variance components.
Regression analysis, Analysis of covariance.

Regression Analysis

Building a regression model: Transformations – Box-Cox model, Stepwise regression, Model selection (adjusted R2, cross validation and Mallow's Cp criteria, AIC and BIC), Multicollinearity.
Detection of outliers and influential observations: residuals and leverages, DFBETA, DFFIT and Cook's Distance.
Checking for normality: Q-Q plots, Normal Probability plot, Shapiro-Wilks test.
Departures from the Gauss-Markov set-up: Heteroscedasticity and Autocorrelation – detection and remedies.
Longitudinal Data Analysis – introduction with motivation. Exploring longitudinal data. Linear models for longitudinal data –introduction, mean models, covariance models, mixed effects models. Predictions.

Types of data. Two-way classified data – Contingency Tables and associated distributions, Types of studies, Relative Risk and Odds Ratio and their properties. More-than-two-way classified data – partial associations, marginal and conditional odds.
Generalized Linear Models: Introduction, Components of a GLM, Goodness of fit – deviance, Residuals, Maximum likelihood estimation.
Binary data and Count data: ungrouped and grouped. Polytomous data.
Over dispersion, Quasi-likelihood.
Models with constant coefficient of variation, joint modeling of mean and variance, Generalized additive models.
Discrete longitudinal data - generalized linear marginal models, GEE for marginal models, Generalized linear subject specific models and transition models.

Design of Experiments

Block Designs: Connectedness, Orthogonality, Balance and Efficiency; Resolvable designs; Properties of BIB designs, Designs derived from BIB designs.
Intrablock analysis of BIB, Lattice and PBIB designs, Row column designs, Youden Square designs; Recovery of inter-block information in BIB designs; Missing plot technique.
Construction of mutually orthogonal Latin Squares (MOLS); Construction of BIB designs through MOLS and Bose's fundamental method of differences.
Factorial designs: Analysis, confounding and balancing in symmetric factorials.

Sample Surveys

Probability sampling from a finite population – notions of sampling design, sampling scheme, inclusion probabilites, Horvitz-Thompson estimator of a population total. Basic sampling schemes – simple random sampling with and without replacement, unequal probability sampling with and without replacement, systematic sampling. Related estimators of population total/mean, their variances and variance estimators – mean per distinct unit in simple random with replacement sampling, Hansen-Hurwitz estimator in unequal probability sampling with replacement, Des Raj and Murthy's estimator (for sample of size two)

in unequal probability sampling without replacement.Stratified sampling – Allocation problem and construction of strata. Ratio, product, difference and regression estimators. Unbiased ratio estimators – probability proportional to aggregate size sampling, Hartley – Ross estimator in simple random sampling. sampling and sub-sampling of clusters. Two-stage sampling with equal/unequal number of second stage units and simple random sampling without replacement / unequal probability sampling with replacement at first stage, ratio estimation in two-stage sampling. Double sampling for stratification. Double sampling ratio and regression estimators. Sampling on successive occasions.

Bayesian Analysis

Different priors and related posteriors

Estimation, testing and prediction for univariate normal distribution with known/unknown mean and/or variance.

Hierarchical and Empirical Bayes under normal setup.

Prior and posterior analysis in generalized linear models

Decision Theory

Risk function, Admissibility of decision rules, Complete, essentially complete, minimal complete and minimal essentially complete classes. Essential completeness and completeness of class of rules based on sufficient statistic and the class of nonrandomized rules for convex loss

Resampling Techniques

Empirical distribution function and its properties

Jackknife and Bootstrap procedures for estimating bias and standard error.

Consistency of the Jackknife variance estimate in an iid setup.

Bootstrap confidence intervals.

Stochastic Processes

Poisson process. Renewal Theory: renewal processes, renewal function, elementary renewal theorem, applications, Blackwell's theorem and key renewal theorem (statements), applications, alternating renewal processes, applications to limiting excess and age.

Markov chains: time-homogeneity, one-step & multi-step transition probabilities, Chapman-Kolmogorov equations, Markov times, strong Markov property, classification of states, stationary distributions, periodicity, ergodicity, convergence, convergence rate. Examples: birth-and-death processes, branching processes.

Jump-Markov processes: conservativeness, transition probabilities, holding times, embedded Markov chain, Chapman-Kolmogorov equations, Kolmogorov backward and forward equations, stationary distributions. Examples: pure birth, birth-and-death chains, Markovian queues.

Time Series Analysis

Stationary time series. Autocorrelation and partial autocorrelation functions. Correlogram. Box-Jenkins Models – identification, estimation and diagnostic checking.

Volatility – ARCH, GARCH models.

Multivariate Analysis:

Multivariate normal distribution and its properties- marginal and conditional distributions. Random sampling from a multivariate normal distribution- UMVUE and MLE of parameters, joint distribution of sample mean vector and SS-SP matrix; Wishart distribution and its properties. Distribution of sample correlation coefficients, partial and multiple correlation coefficients, partial regression coefficient and intraclass correlation coefficient. Distributions of Hotelling's T^2 and Mahalanobis' D^2 statistics- their applications in testing and confidence set construction. Multivariate linear model, MANOVA for one-way and two-way classified data.

Applied Multivariate Analysis

Clustering: Hierarchical clustering for continuous and categorical data- different choices of proximity measures, Agglomerative and divisive algorithms.K-means clustering- optimum choice of the number of

clusters.

Classification and discrimination procedures: Discrimination between two known populations – Bayes, Minimax and Likelihood Ratio procedures. Discrimination between two multivariate normal populations. Sample discriminant function. Likelihood ratio rule. Tests associated with discriminant function, Probabilities of misclassification and their estimation. Classification of several populations. Fisher's method for discriminating among several populations.

Principal Component Analysis: Population and sample Principal components and their uses. Plotting techniques, Large sample inferences.

Factor Analysis: The orthogonal factor model, Estimation of factor loading, Factor rotation, Estimation of Factor scores, Interpretation of Factor Analysis.

Canonical Correlations: Population and sample canonical variables and canonical correlations and their interpretations. Plotting techniques, Large sample inferences.