Abstract Booklet

Advances in Statistical Science: International Conference in Celebration of 90th Birth Anniversary of Professor C.R. Rao

AISS 2010

Indian Statistical Institute
Kolkata, India

January 10--11, 2010

Sponsored by:
Ministry of Statistics & Programme Implementation, Government of India and Kolkata Port Trust
Advances in Statistical Science:
International Conference in Celebration of
90th Birth Anniversary of Professor C.R. Rao
AISS 2010

Abstracts

- Abstracts of C.R. Rao Visionary Lectures 3
- Abstracts of Invited Talks 5
- Abstracts of Contributed Talks 16

Note: Within each category, the abstracts are put in alphabetical order based on the last name of the presenter. In case of multiple authors, the presenter of the paper is indicated with an asterisk.
Bayesian approaches are widely used to capture overdispersion in binomial and multinomial data through the use of beta-binomial and Dirichlet-multinomial models. These approaches can be extended to multiple hierarchical levels to produce a partitioning of diversity within- and among-levels in binomial and multinomial models. In this paper, we consider situations in which the prior distribution of a parameter vector in the distribution of the observable binomial and multinomial data contains a hyperparameter vector, which itself has a hyperdistribution and so forth. We first develop a novel multilevel product partitioning result that shows how the total diversity is partitioned among various hierarchical levels. Then we establish that the gain in information decreases as one moves to higher levels of hierarchy. We also establish the connection with the clumped multinomial data. Finally, we apply our methodology to simple multilocus, two-allele genetic models to show how diversity is partitioned into diversity within populations, diversity among populations within geographical regions, diversity among geographical regions, and diversity among loci.

Statistical Challenges and Opportunities in the Biomedical Sciences

Robert C. Elston
Case Western Reserve University, Cleveland OH, USA

My earliest statistical work was motivated by some of C.R. Rao’s seminal papers and books; our common interests have included the estimation time-response curves, variance components and unbiased prediction. Risk prediction requires an understanding of probability. Frequentist and Bayesian approaches are compared in terms of what the scientist wants, and it is clear that in modern genomic research the latter is possible without any of what Fisher called the ”Bayesian postulates”. Finally I recall things I have learned from working with biomedical colleagues and hearing talks given by other statisticians, and how these should relate to the way we teach statistics.
A generalized censoring scheme in the survival analysis context was introduced by Jammalamadaka and Mangalam (2003, Jour. of Nonparametric Statistics, pp.253-265). In this talk we discuss how such a censoring scheme applies to circular data and in particular when the original data is assumed to come from a parametric model such as the von Mises. Maximum likelihood estimation of the parameters as well as their large sample properties are considered under this censoring scheme. Monte Carlo comparisons are made with alternate estimators of the mean direction and concentration.
Invited Talks

Micro-Macro Linkages as Statistical Framework for Economic Analysis and Policy

R.B. Burman
Reserve Bank of India

The System of National Accounts (SNA), now 2008 version, and related systems for balance of payments, fiscal and financial statistics form the basic building blocks of a country’s statistical system. These statistics have served the needs for the evaluation of the economy on issues such as economic growth, saving and investment, internal and external balance, economic stabilization policy options etc. very well. However, volatility, instability and macroeconomic imbalances continue to challenge researchers, and often, they are confronted with surprises and puzzles, such as Greenspan conundrum. Sir Richard Stone, the main architect of the SNA and Richard Ruggles strongly felt the need of micro-macro linkages as building blocks of SNA for shedding much more light on the dynamics of the economy. There has been progress on this front but challenges are many for developing a workable consistent framework for such a statistical system for a country. The presentation will shed some light on developing such a system in stages, given the present state of information technology and preparedness. Needless to say, we need such a framework in place early to pursue economic policy which will not only focus on competitive market mechanism supporting productive initiative but also be socially optimal.

Family-Based Association Studies for Gene Mapping

Varghese George
Medical College of Georgia, Augusta, GA, USA

Linkage and association analyses are the two main approaches for mapping disease genes and markers underlying complex diseases. Both population-based and family-based designs are commonly used in genetic association studies. The family-based designs are robust against population admixture and stratification, and thus allow testing for association due to linkage disequilibrium. The simplest family-based design for association is the transmission disequilibrium test (TDT) design for dichotomous traits. In this presentation, we explore some of the extensions of TDT for both dichotomous and quantitative disease traits under more generalized sampling schemes.
Fractile Graphical Analysis and Non-Parametric Regression in a New Perspective

Anil Bera and Aurobindo Ghosh*
University of Illinois, IL, USA

Fractile graphical analysis (FRG) was proposed by Prashanta Chandra Mahalanobis in a series of papers (e.g., Econometrica, 1960) and seminars as a method for comparing two distributions controlling for the rank of covariate through fractile groups. Mahalanobis used a heuristic method for approximating the standard error of the dependent variable using fractile graphs from two independently selected “interpenetrating subsamples.” First we revisit the technique of FGA with some historical perspective. Then we evaluate a new non-parametric regression method called “fractile regression,” where we condition on the rank of the covariate, and compare it with existing regression techniques. Finally, we apply this method to compare mutual fund inflow distributions conditional on returns.

Understanding the Transportational and Depositional Setting of Panchet Formation, Purulia and Bankura Districts, West Bengal from Grain Size Analysis

Dipsikha Ghosh1*, Bapi Goswami1 and Abhik Kundu2
1. University of Calcutta, Kolkata, India 2. Asutosh College, Kolkata, India

The rocks of Panchet Formation, of Upper Gondwana age predominantly include sandstone of medium to fine grain-size, with subordinate shale and pebble horizon cemented by carbonate cement (mostly calcitic). Sediment grains sizes and the statistical parameter like mean, standard deviation, skewness and kurtosis are used to understand the transportation, and the depositional environment of the sediments in a part of Panchet Formation. Granulometric analyses indicate the presence of fluvial environment interrupted with aeolian, beach and tidal environments prevailed during the time of deposition of the sediments.

Designing Exons for Human Olfactory Receptor Gene Subfamilies Using a Mathematical Paradigm

SK Sarif Hassan, Pabita Pal Chowdhury, Amita Pal, R.L. Brahmachary and Arunava Goswami*
Indian Statistical Institute, Kolkata, India

Ligands for only two human olfactory receptors are known. One of them, OR1D2, binds to Bourgeonal, a volatile chemical constituent of the fragrance of mythical flower, Lily of the valley.
or Our Lady’s tears, Convallaria majalis (also the national flower of Finland) [Malnic B, Godfrey P-A, Buck L-B (2004) The human olfactory receptor gene family. Proc. Natl. Acad. Sci U. S. A. 101: 2584-2589 and Erratum in: Proc Natl Acad Sci U. S. A. (2004) 101: 7205]. OR1D2, OR1D4 and OR1D5 are three full length olfactory receptors present in an olfactory locus in human genome. These receptors are more than 80 identical in DNA sequences and have 108 base pair mismatches among them. Apparently, these mismatch positions show no striking pattern using computer pattern recognition tools. In an attempt to find a mathematical rule in those mismatches, we find that L-system generated sequence can be inserted into the OR1D2 subfamily specific star model and novel full length olfactory receptors can be generated. This remarkable mathematical principle could be utilized for making new subfamily OR members from any OR subfamily. Aroma and electronic nose industry might utilize this rule in future.

The use of circular statistics to detect lunar synchronisation in zoology, and a case study of amphibian arrivals at three breeding sites

Rachel Grant
Oxford University, UK

Circular statistics are widely used in the biological sciences, but their use is often limited to studies of direction, for example orientation behaviour. Many marine and nocturnal terrestrial animals synchronise certain aspects of their behaviour to the lunar or semi-lunar (tidal) cycles. Although periodicity of this type can be analysed by time series analysis, in many biological situations the data to be analysed is not suitable for this type of analysis. Hence, lunar periodicity in zoology have been traditionally analysed in a variety of ways. A survey of recent literature on the subject showed that in many cases the lunar cycle is merely divided into arbitrary phases and a chi squared test for goodness of fit is used. However, because the divisions are arbitrary, using different divisions will result in different statistical outcomes. Therefore there is a high possibility of both type 1 and type 2 errors. Linear regression has often been used to detect lunar periodicity but the lunar cycle is a circular, not a linear variable, so linear regression is invalid in this case as the origin is arbitrary or undefined. However, the use of linear regression has persisted in studies of this phenomenon as biologists are generally unaware of the possibility of using circular regression models.

In our study of amphibian reproduction we monitored arriving and mating animals over several breeding seasons at three breeding sites in Europe. For preliminary analysis we looked at the frequency of large arrival, mating and spawning events and used Rao’s spacing test for departure form randomness to test for the influence of the lunar cycle. However, although this showed that some amphibian species synchronise their reproduction with the full moon, it did not allow us to simultaneously analyse other environmental variables that could have affected the amphibians. It is often necessary when studying amphibians in particular to take into account the influence of weather variables as amphibian behaviour is highly influenced by rainfall and temperature.
Weather data of this type are invariably linear and there is a need to carry out a simultaneous multivariate analysis that is capable of analysing linear and circular variables together in order to understand the relative influences of lunar phase and weather. We also wanted to look at the influence of the earth’s magnetic field, represented by Kp.

As the data are count data, they follow a Poissonian distribution. There is also an issue of over dispersion in the data, as there are many days when no amphibians arrive resulting in many zero values. We wanted to model the response of both male and female amphibians arriving. Hence we used maximum likelihood estimation to fit a diagonally inflated bivariate poisson model using the following regressors: maximum and minimum temperatures, \( \cos(Kp) \), \( \sin(Kp) \), \( \cos(\text{days since the last full moon}) \), \( \sin(\text{days since the last full moon}) \). This analysis showed that the lunar cycle does influence the arrival rates of amphibians at the breeding site but that rainfall, traditionally important in regulating amphibian movements, is not a significant predictor of arrival rates for this data set. This is probably because one particular species (the common toad Bufo bufo) was highly represented in the dataset. This is a terrestrial toad which may be better at conserving water than more aquatic species. In order to test this we will carry out separate analysis by species and breeding site to look for differences in response to the lunar cycle and weather.

A Statistical Approach To Study The Maximum Drilling Depth For Groundwater Production Wells

U.de S. Jayawardena
University of Peradeniya,Sri Lanka

Hydrogeological investigations were carried out to provide high quality groundwater for the people living in rural areas. Under this programme 14 production wells were constructed by drilling to use available groundwater at the preselected locations where the elevations are comparatively lower than the other areas in the studied area.

The borehole reports in which the detailed information of drilled locations have been recorded, were used to analyze the subsurface fracture systems and the variations of flushing yields with depth and relevant mean sea levels. The total flushing yields and total number of fractures in each 5 meter depth of every well were used as frequency tables.

The results indicate that drilling more than 60 meters below the surface level is a loss and no sufficient groundwater will be available except in a very few locations. A maximum drilling depth according to mean sea level cannot be recommended because horizontal fractures may not occur in a complex geological area. These results are valid only for this region but the introduced method can be used for any other area to find a maximum economical drilling depth.
Wald Tests for Detecting Multiple Structural Changes in Persistence

Mohitosh Kejriwal
Purdue University, USA

This paper considers the problem of testing for multiple structural changes in the persistence of a univariate time series. We propose sup-Wald tests of the null hypothesis that the process has an autoregressive unit root against the alternative hypothesis that the process alternates between stationary and unit root regimes. Both non-trending and trending cases are analyzed. We derive the limit distributions of the tests under the null and establish their consistency under the relevant alternatives. The computation of the test statistics as well as asymptotic critical values is facilitated by the dynamic programming algorithm proposed in Perron and Qu (2006) which allows the minimization of the sum of squared residuals under the alternative hypothesis while imposing within and cross regime restrictions on the parameters. Finally, we present Monte Carlo evidence to show that the proposed tests perform quite well in finite samples relative to those available in the literature. This is joint work with Pierre Perron and Jing Zhou.

On the Power of Certain Panel Unit Roots Tests under Reasonable Alternatives

Kajal Lahiri* and Xiaomei Li
University at Albany, USA

This paper explores the power characteristics of a number of unit root tests in dynamic heterogeneous panels. It is argued that the specification of the alternative hypothesis in most studies (cf, IPS test in Im, Pesaran and Shin, Journal of Econometrics, 2003) is too restrictive and implausible in varied empirical settings. In reality, under a reasonable alternative, only a fraction of the cross section units will be stationary with widely different values of the autoregressive coefficients, and the stationary subset is seldom known a priori. In this study, AR coefficients are generated from Weibull and Generalized Exponential distributions right-censored at one. We also allow for cross-section correlation in errors and consider Fisher-type p-value test popularized by Maddala and Wu (Oxford Bulletin1999). Results show that the power does not necessarily increase with the number or fraction of stationary units in the panel; other characteristics of the distribution of the AR parameter are also important. The Fisher-type tests out perform IPS-type tests when the panel contains more unit roots, but is more sensitive to the variation in the AR coefficient distribution. In particular, if the whole panel is stationary, the Fisher-type tests become less powerful than the IPS-type tests, with and without cross section dependence.
Multi-objective Optimal Experimental Designs in Event-Related fMRI Studies

Abhyuday Mandal
University of Georgia, Athens, USA

Event-related functional magnetic resonance imaging (ER-fMRI) is considered one of the leading technologies for studying human brain activity in response to mental stimuli. With sophisticated allocations of stimuli, researchers can gather valuable fMRI time series and acquire precise information about human brain activity. However, due to the nature of fMRI experiments, the underlying design space is very large and irregular. This makes it difficult to find an optimal design that simultaneously accomplishes various goals of a study and fulfills the scientific restrictions. Kao, Mandal, Lazar and Stufken, 2009, NeuroImage, 44: 849-856, propose an efficient approach for finding good ER-fMRI designs. Here we consider multiple objectives, including estimating the hemodynamic response function (HRF), detecting activation, circumventing psychological confounds and fulfilling customized requirements. Taking into account these goals, we formulate a family of multi-objective design criteria and develop a genetic-algorithm-based technique to search for optimal designs. Our proposed technique incorporates existing knowledge about the performance of fMRI designs and its usefulness is shown through simulations. As other methods known hitherto, this approach also fails to take into account the link between two common statistical goals, namely, the estimation of the hemodynamic response function and the detection of active brain voxels. To overcome this, we consider a nonlinear model to accommodate both goals in a unified setting. Under this model, we define design criteria to evaluate designs and search for local and Bayesian optimal designs via a knowledge-based genetic algorithm. We also compare designs that we obtain to those currently in use by researchers. (This talk is based on joint research with Ming-Hung (Jason) Kao, Dibyen Majumdar and John Stufken).

Optimal Allocation for Multi-level Stress Testing with Weibull Regression Under Censoring

Hon Keung Tony Ng*, C.Y. Ka, P. S. Chan and N. Balakrishnan
Southern Methodist University, USA

Optimal allocation problem in a multi-level stress test with censoring and Weibull (extreme value) regression model is discussed. The maximum likelihood estimators and their Fisher information and asymptotic variance-covariance matrix are derived. Three optimality criteria are used to discuss the optimal allocation problem. Optimal allocation of units, both exactly for small sample sizes and asymptotically for large sample sizes, for two- and four-stress-level situations are determined numerically. Conclusions and discussions are provided based on the numerical studies.
Cluster Analysis: Applications in Life Science Problems

M. Pandey
Banaras Hindu University, Varanasi, India

Often in a scientific study of a real life or a research oriented problem, the necessary data is multidimensional. A suitable multivariate technique is required to meet the objective of the study.

This paper considers applications of cluster analysis in two problems of Life Science. One relates to the episodes of the water borne enteric (WBE) diseases in Varanasi city and the other is related to the 16 ecotypes of Hydra from 15 locations all over India.

For the first study a survey was undertaken in the OPD of the four reputed hospitals of the Varanasi city. Each of the hospital was visited once a month for a duration of one year from July 1998 to June 1999. Total 343 outdoor patients of following five WBE diseases (i) Diarrhoea (ii) Dysentery (iii) Hepatitis (iv) Typhoid and (v) Gastro-enteritis were interviewed with a structured questionnaire. All possible answer types were coded and given different scores according to their vulnerability towards enteric diseases. In all 24 variables have been used to identify, if there exists an inherent grouping among the said OPD patients. The analysis resulted in only five clusters.

In the second problem 35 variables at various levels of organization have been considered. 6 to 7 clusters emerged for various levels and for all levels combined. Further a stepwise discriminant analysis was applied to see if smaller number variables are sufficient for proper classification of Hydra in eco-groups. It has been found that only 6 variables are sufficient to define the classification functions, which are capable of 100

Gender Difference in Employment Among Different Religious Groups in India

R. N. Pandey
Ministry of Tourism, Govt. of India, New Delhi, India

India has lot of socio-economic, demographic, geographic and cultural diversities. In spite of the implementation of various policies and programmes for either elimination or reduction of various discriminations, some kinds of discrimination are still reported. The extent of involvement of women in unpaid work is higher than that of men in almost all religions and ethnic groups. Using Work Participation Rate (WPR) for women, in this paper an attempt has been made to study the gender difference in WPR among various religious groups in India and its trend over the period using the data from the Employment and Unemployment Surveys conducted by the National Sample Survey Organisation (NSSO). Analysis attempted in this paper clearly brings out the gender difference in employment situation in India. Though, this difference persists among all the
religious group, quantum varies. The WPRs for Christian women was generally found to higher followed by Hindus and Muslims.

Changing Patterns Of Arsenic Concentration Around Hotspots In The Deltaic Alluvial Plain Of West Bengal, India

Barendra Purkait¹*, Ashis Sengupta² and Moumita Roy²
1. Publication Division, Geological Survey of India, Kolkata, India 2. Applied Statistics Unit, Indian Statistical Institute, Kolkata, India

A total of 700 drinking water samples were collected from tube wells covering an area of about 1000 sq km of the deltaic alluvial plain of the Malda district of West Bengal, India. Arsenic (As) content of water was measured during field survey by E-Merck Arsenic field kit with some laboratory replications. Arsenic hotspots (As > 0.01 mg/l as per WHOs permissible limit and As > 0.05 mg/l as per Indian Standard) were detected at different blocks. Different manifestation of arsenic poisoning was observed among the people of the area. As-content in the water varies in wide limit in the same depth even within a radius of 15m. Thus the arsenic distribution pattern with its migration trend in the surrounding areas around the hotspots requires thorough investigation. From the plots of As-content of water, arsenic zonation map was prepared with six groups: As 0−0.01mg/l; 0.01−0.02mg/l, 0.02−0.03mg/l, 0.03−0.04mg/l, 0.04−0.05mg/l and > 0.05mg/l. With reference to an As-sink area (Kaliachak police station), Arsenic distribution pattern was delineated and the circular histograms were drawn with corresponding vector means of the distribution trend. Statistical test was conducted for the test of uniformity whether the distribution pattern is isotropic or anisotropic. Possible explanation was given on the causative factors for such distribution pattern. The results might be of helpful in the genesis of arsenic vis−vis its contamination pattern in the surrounding areas around the hotspots.

Bivariate Lack Of Memory Property

Dilip Roy
The University Of Burdwan, West Bengal, India

Marshall and Olkin (1969) introduced a bivariate generalisation of bivariate lack of memory property. That generalisation under marginal exponential distributions has led to characterisation of the family of bivariate exponential distribution due to Marshall and Olkin.

Unfortunately, the lack of memory property proposed by Marshall and Olkin cannot be reduced to univariate lack of memory property. In view of this limitation an alternative definition for bivariate lack of memory property will be examined which reduces to univariate lack of memory property and characterized the bivariate exponential distribution due to gumbell (1960). A few
other notes related results will be highlighted along with a possible extension in the stochastic domain.

*Mathematical Methods of Statistics*

Frechet-differentiation of functions of operators with application to testing the equality of two covariance operators

Frits H Ruymgaart and Krishna Kaphle*
Texas Tech University, Lubbock, USA

It is well known that the sample covariance operator converges in distribution in the Hilbert space of Hilbert-Schmidt operators, and that this result entails the asymptotic distribution of simple eigenvalues and corresponding eigenvectors. Several estimators and test statistic for the analysis of functional data require the asymptotic distribution of eigenvalues and eigenvectors of certain functions of sample covariance operators. Because the sample covariance operator can not be one-to-one in infinite dimensional Hilbert spaces, important examples of such functions are related to a generalized inverse (cf: C.R. Rao and P.K. Mitra (1971): *Generalized inverse of matrices and its applications*, Wiley, New York.) To obtain such a result, it turns out that the asymptotic distribution of such a function of the sample covariance operator is a prerequisite. We briefly review the Fréchet-derivative of functions of operators and an ensuing delta-method to solve this problem. The results are applied to obtain the asymptotic distribution of a statistic for testing the equality of two covariance operators. The test with restricted null hypothesis is already in literature. Here, we use a different method where this restriction is not needed.

*Modeling Data from High-throughput ChIP-Seq Experiments*

Deepayan Sarkar
Indian Statistical Institute, Delhi, India

The advent of high-throughput sequencing of short DNA fragments has opened up the opportunity to study DNA-protein interactions in great detail through ChIP-Seq experiments. The premise of ChIP-Seq experiments is conceptually simple: DNA fragments in direct physical interaction with transcription factors are isolated by chromatin immunoprecipitation (ChIP), and then partially sequenced. However, there are many details that need to be addressed in order to ensure that the data obtained from ChIP-Seq experiments are correctly interpreted. In this talk, I will describe a natural null model for the background noise in ChIP-Seq data and propose simple methods for estimation of the null model and hypothesis testing when the data contains both noise and signal. If time permits, I will also discuss the important but difficult-to-formulate problem of sample-size (power) calculations.
On Bootstrap Based Testing for Jumps Under Long-range Dependence

Jan Beran and Yevgen Shumeyko∗
University of Konstanz, Germany

The aim of this paper is to construct an appropriate statistic for testing the occurrence of discontinuities in a trend function when the residual process exhibits long memory. It is shown that the test statistic based on a moving block bootstrap and a wavelet decomposition is asymptotically normal with smaller variance under the null hypothesis of a continuous trend function.

A Systematic Approach For Unequal Allocations Under Ranked Set Sampling With Skew Distributions

Neeraj Tiwari∗ and Girish Chandra
Kumaon University, Almora, India

Ranked Set Sampling (RSS) is a useful technique for improving the estimates of mean and variance when the sampling units in a study can be more easily ranked than actually measured. Under equal allocation, RSS is found to be more precise than simple random sampling (SRS). Further gain in precision of the estimate may be obtained with appropriate use of unequal allocation. For skewed distributions, the optimum gain in precision is obtained through unequal allocation based on Neymans approach, in which the sample size corresponding to each rank order is proportional to its standard deviation. However, the unavailability of the standard deviations of the rank orders makes the Neymans approach impractical. The two models, viz., t-model and (s, t)-model suggested by Kaur, Patil and Taillie (1997) are also impractical due to their dependence on population parameters of rank orders and complexities in finding the optimum values of t and (s, t). In this article, we propose a simple and systematic approach for unequal allocation for RSS with skew distributions. The proposed approach performs better than SRS and RSS with equal allocation. It also appears to perform better than the RSS with unequal allocation using t-model and quite close to the (s, t)-model in most of the situations we have considered. The performance of the proposed procedure relative to existing models has been numerically evaluated for some skewed distributions.

Variance Bounds for Estimators in Autoregressive Models with Constraints

Anton Schick1, Ursula U. Müller2 and Wolfgang Wefelmeyer3∗
1. Texas A&M University 2. Binghamton University 3. Universität zu Köln

We consider nonlinear and heteroscedastic autoregressive models whose residuals are martingale increments with conditional distributions that fulfill certain constraints. We treat two classes
of constraints: residuals depending on the past through some function of the past observations only, and residuals that are invariant under some finite group of transformations. We determine the efficient influence function for estimators of the autoregressive parameter in such models, calculate variance bounds, discuss information gains, and suggest how to construct efficient estimators. Without constraints, efficient estimators can be given by weighted least squares estimators. With the constraints considered here, efficient estimators are obtained differently, as one-step improvements of some initial estimator, similarly as in autoregressive models with independent increments.
Contributed Talks

The Error Distributions for cDNA Microarray Gene Expression Data

P.P.Bindu\(^1\), Sangita Kulathinal\(^2\) and Sebastain George\(^1\)
\(^1\) St.Thomas College, Pala, Kerala, India \(^2\) The Indic Society for Education and Development, Nashik, India

Microarrays have become an important tool for studying the molecular basis of complex disease traits and fundamental biological processes. A common purpose of microarray experiments is the detection of genes that are differentially expressed under two conditions such as treatment versus control. In microarray, the distribution of gene expressions usually exhibits heavier tails than a Gaussian distribution and has a certain degree of asymmetry. Therefore, this distribution has been modeled in the literature using different parametric families of distributions, such the Asymmetric Laplace or the Cauchy distribution. Moreover, a power law describes the tails of spot-intensity distributions and the variance of a given array increases with the number of genes. In the present study we developed an asymmetric slash Laplace distribution and skew-double Lomax distribution as the error distribution for gene expression data. These two models exhibit heavier tails and skewness of varying degrees. Hence these models provide the flexibility for modelling impulsiveness and skewness required for the gene expression data. We have explored various properties and estimated the parameters of these models. Finally, possible applications of the models are highlighted.

Tschuprow-Neyman Optimum Allocation Under Superpopulation

B.K.Gupt
North Eastern Hill University, Shillong, India

In stratified sampling one of the major problems is the allocation of total sample size to the strata. Tschuprow (1923) Neyman (1934) Optimum Allocation [TNOA] involves unknown population parameters for which at least proportionate values are required. In practice, corresponding functions of a closely related auxiliary variable are substituted in place of the unknown parameters for computation of the TNOA. The justification for this substitution was provided for the usual expansion estimator under superpopulation model (SPM) [Cochran (1946)] approach first by Hanurav (1965) and then by Rao T.J. (1968,77) who suggested allocations based on the auxiliary variable for stratified simple random sampling with and without replacement (SRSWR and SRSWOR) designs. Gupt (2003) considered a more general correlated SPM and obtained to generalized allocations of which Hanurav (1965) and Rao (1968) allocations follow as particular
In this paper the problem of allocation of sample size to strata is studied for stratified SRSWR design under the usual SPM considered by Rao (1977) in which 

$$V(Y_{ij} | X_{ij}) \propto X_{ij}^g$$

where $$Y_{ij}, X_{ij}$$ are respectively the values of the study and auxiliary variables and $$g$$ is a SPM parameter. Expectation of the unknown parameters are derived under the model and the two Model Optimum Allocations (MOA) are obtained as follows: 1. $$n_i \propto N_i \sigma_i(x)$$ and 2. $$n_i \propto (N_i \sum X_{ij}^g)^{1/2}$$ provided $$\theta_i(g) = \sigma_i(x)/[(1/N_i) \sum X_{ij}^g] \propto \text{Constant}$$. Further, two allocations under the assumption of equal stratum coefficient of variation (ESCV) viz., Mean-based, MESCVMOA. $$n_i(g) \propto \sum X_{ij}^{g/2}$$ and SD based SDESCVMOA $$n_i(g) \propto N_i \sigma_i(x^{g/2})$$ are also derived under appropriate conditions. Furthermore approximate allocations from all the model based allocations are obtained under the large strata approximation. Finally the results are illustrated with two live populations. We observe as the stratum sizes are usually large in survey population in practice, the large stratum approximation does not have any substantial effect on the efficiency of an allocation. On the basis of 2 live populations, we note that either of the ESCV allocation is better than MOA from which they were derived the choice between the two ESCV assumption based MOA may be made by comparing the condition CV of the two, the lower the CV, the higher efficiency may be expected. Thus the level of satisfaction of the condition is helpful in choosing the best possible allocation for use in practice at the planning stage of the survey.

Small And Efficient Designs Of Constrained Mixture Problem For Fitting Quadratic Model

Chhaya Pramod Jadhav$^{1*}$ and Jyoti Rajarajan$^{2}$

$^1$Dr. Babasaheb Ambedkar University Aurnagabad, M.S., India, $^2$Sardar Patel University,Vallabh Vidyanagar, Gujarat, India

Experiments that involve the blending of several components are known as mixture experiments. In some constrained mixture experiments, even a small change in proportions of one or more components, produces a curvilinear change in response behavior. To model such a change would require the use of at least a second degree equation. Saxena and Nigam (1977) have given symmetric simplex design for fitting second order models for constrained region. Though symmetric simplex design is having less no of design points, it doesn’t guarantees about the efficiency of the design. G. F. Pipel in (1988) presented the CONVERT and CONAEV algorithms to obtain the extreme vertices, mid points of edges and face centroids to fit a quadratic model for constrained regions. Pipels program also generates the designs however it is with large no of design points and hence its impractical to use.

Here we propose a method for design of constrained mixture experiments to fit quadratic model which is small and at the same time efficient also.
Construction of Optimal Mixed-Level Supersaturated Designs

Basudev Kole
IASRI, Pusa, New Delhi, India

This article describes some methods of construction of mixed level $f_{NOD}$-optimal supersaturated designs. The methods of construction exploit the layout and the property of Uniform designs and Hadamard matrices. Mathematical expression for $E(f_{NOD})$ and $E(\chi^2)$ criteria have been obtained for many designs constructed in this paper. Some examples are given to illustrate the methods of construction. A catalogue of 67 optimal mixed level supersaturated designs with at most 60 runs and 60 factors is prepared. Some other important features of the designs are also given in the catalogue. All designs are $f_{NOD}$-optimal while some designs are $\chi^2$-optimal too.

Reliability Analysis Of a Two-Unit Parallel System With Inspection and Priority Subject to Degradation

M.S.Kadyan$^2$, S.C.Malik$^1$ and Jitender Kumar$^2$*
1. M. D.University, Rohtak, India 2. Kurukshetra University, Kurukshetra, India

This paper deals with the stochastic behavior of a two-unit parallel system in which unit becomes degraded after repair and a single server who attend the system immediately whenever needed. The server inspects the degraded unit at its failure to see the feasibility of repair. If the repair of the degraded unit is not feasible, it is replaced by new unit which gets priority in repair over the degraded unit. The system is considered in up-state if any of new/degraded unit is operative. The time of the failure of the units are negative exponential distributed while that of inspection and repair times arbitrarily distributed. The system is analyzed by using semi-Markov process and regenerative point technique to obtain various and economics related measures. A particular case is considered to depict the behavior of some characteristics of interest graphically.

On Estimation of Growth Rates Using Linear and Exponential Functions

Sananse, S and Maidapwad S. L$^*$
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India

The growth rates are being estimated mostly through the linear or exponential functions. A number of research publications have used this procedure in various situations and estimated growth rates which is being used by planners or policy makers. However, the data may not be following these linear or exponential models or may require fitting of higher degree polynomials or non-linear models. Further, these models lack the econometric considerations i.e. autocorrelations, homoscedasticity and randomness. As such, it is necessary to make a critical review of the existing methodologies used for estimation of growth rates and to propose appropriate alternatives.
The present study is an attempt by way of putting forth the methodology for the computation of the growth rates based on these statistical criteria as well as the econometric criteria. This proposed methodology is based on the econometric considerations satisfying the assumptions of the Ordinary Least Squares Method about the random error term. The proposed procedure can be used to estimate accurate and precise growth rates in most of the situations.

Construction of Neighbour Balanced BIBRC Designs

Anurup Majumder∗ and Aatish Kumar Sahu
Bidhan Chandra Krishi Viswavidyalaya, West Bengal, India

A new method of construction of a series of BIBRC \((v, b, r, p, q, \lambda)\) designs, where \(v\) is a prime or prime power has been presented in this paper. The series of BIBRC designs is developed from mutually orthogonal mates (MOMs) of BIBRC designs. The BIBRC designs constructed by using the above mentioned method are found to be neighbour balanced as they satisfy the condition that every pair of treatments occurs in neighbouring plots \(n_r\) times in rows, \(n_c\) times in columns and \(n_d\) times in diagonals. Available literature reveals that the developed neighbour balanced BIBRC designs are more flexible to \(p\) and \(q\) for a particular value of \(v\). A new series of BIBRC designs has also been developed for even number of \(v(= s, \text{where } s − 1 \text{ is prime})\) from orthogonal mates (OM) of BIB designs.

The Historical Development Of Robust Estimators

Muthukrishnan.R∗ and M.Radha
Bharathiar University Coimbatore,Tamilnadu, India.

In statistics, the classical (traditional) procedures should perform reasonably well if certain assumptions hold, but may be unreliable if one or more of these assumptions are violated. A robust analog of a given classical procedure should also work well when these assumptions hold, but the robust procedure is generally tailored also to give useful results when a single, specific assumption is relaxed. Robust Statistical methods were first introduced in computer vision to improve the performance of feature extraction algorithms. These methods can tolerate the presence of data points that do not obey the assumed model (up to 50%). However, they can break down at unexpectedly lower percentages when the outliers are clustered. The traditional robust methods generally assume that the data of interests occupy a majority of the whole data. In image analysis, however, the data is often complex and several instances of a model are simultaneously present, each accounting for a relatively small percentage of the data points. To deal with data including multiple structures and a high percentage of outliers (more than 50%) remains a challenging task. In this paper we have reviewed most frequently used robust estimators, since its inception and in the context of computer vision.
The $q$-Laplace Distribution and Its Applications in Financial Modeling

Shanoja R. Naik$^*$ and K.K. Jose

Pala Campus, Arunapuram, Kerala, St. Thomas College, Pala, Kerala, India

In this paper we introduce the $q$-Laplace distribution which facilitates a transition to the Laplace distribution as $q \rightarrow 1$. We study its properties in detail and show that it can be derived by optimizing the Mathais entropy subject to certain constraints. Also we consider the distributions of the product and ratio of two random variables coming from the same family of $q$-Laplace distributions and illustrate the use of the $q$-Laplace distribution in modeling currency exchange rates.

D-Optimal Designs For Three And Four Component Mixture Models With Inverse Terms

Mahesh Panda

University of Delhi, India

In mixture experiments, the measured response is assumed to depend only on the relative proportions of the components present in the mixture. If $x_i$ represents the proportion of the $i^{th}$ component in a $q$ component mixture then

$$0 \leq x_i \leq 1, \quad i = 1, 2, \ldots, q \quad \text{and} \quad \sum_{i=1}^{q} x_i = 1.$$ 

As a result, the experimental region is a $q - 1$ dimensional simplex given by

$$S_{q-1} = \left\{ (x_1, x_2, \ldots, x_q)' : \sum_{i=1}^{q} x_i = 1, \quad x_i \geq 0, \quad 1 \leq i \leq q \right\}.$$ 

Cornell (2002) described a number of model forms for mixture experiments. Draper and St. John (1977a) introduced models with inverse terms to model extreme changes in the response behavior when the value of $x_i$ becomes very close to zero. Extreme changes in the response behavior are very real occurrences in many areas of chemical experimentation, especially when certain component proportions approach boundary conditions.

Chan and Guan (1994) analytically obtained D-optimal designs for models with inverse terms for two mixture components for

$$\Delta(\delta) = (x_1, x_2, \ldots, x_n)' : \quad x_i \geq \delta \quad (i = 1, 2, \ldots, q), \quad \sum_{i=1}^{q} x_i = 1.$$ 

In this paper, we have analytically obtained the $D$-optimal designs for these models in three and four components for the design space $\Delta(\delta)$ where $\delta \in (0, 1/q)$ is arbitrary but fixed.
One Step Method to Calculate the Chi-square Value

Achhelal Pasi
National Institute for Research in Reproductive Health, Mumbai, India

Conventional formula for calculation of chi-square value requires calculation of Expected value (E-Value). This makes the process bit lengthy and time consuming. Present paper will come out the mathematical analysis of existing formula to calculate chi-square value in one step and without calculating Expected value. Application of the method will really save the time on calculation.

A New Test for Testing Significance of Factorial Effects in Unreplicated Factorial Designs

Kulkarni H.V. and Patil S.P.*
Shivaji UNiversity, Kohlapur, India

In many industrial experiments, unreplicated full or fractional factorial designs are a common practice. Usual F-test based analysis procedure is not available for such experiments due to lack of an estimate for the common error variance. A large number of alternative strategies have been suggested in the literature. We propose a new formal test for the analysis of unreplicated factorial experiments based on the orthogonality of the columns of the underlying design matrix, which is a typical property of factorial designs. The response variable is assumed to be normally distributed. The power function of the proposed test is studied and its comparison with the existing methods is our future plan.

A Simple Normal Approximation for Weibull Distribution with Applications to Interval Estimation

H. V. Kulkarni and S. K. Powar*
Shivaji UNiversity, Kohlapur, India

In this article we consider the problems of confidence interval (CI) estimation for Weibull mean and estimation of upper prediction interval that includes at least \( l \) out of \( m \) future observations from a Weibull distribution at each of \( r \) locations based on a proposed simple Normal Approximation. A comparative study revealed that besides being computationally very simple and fast compared to its main competitor, namely, the CI based on Generalized Variable (GV) approach, the proposed CI for Weibull mean have very good coverage probabilities, that are uniformly larger than GV based CI while maintaining almost equivalent expected lengths. The proposed normal approximation based prediction intervals exhibit exactly similar performance as that for interval estimation of mean. The problems considered are illustrated using some real data sets.
A Test for Main Effects When Observations Are Randomly Right Censored

A. Shanubhogue and D.P. Raykundaliya∗
Sardar Patel University, Vallabh Vidyanagar, India

In this article we have generalized the Gore-test (Gore (1972)), along the lines of Breslow (1970) and Patel (1975), for randomized block design when the observations are subject to arbitrary right censorship. The distributions of censoring variables are allowed to vary from block to block. The asymptotic distribution of the proposed statistic is shown to be chi-square. The power of the proposed test statistic is simulated and found to be relatively better than Patel test. The computation is done by taking distribution of error as exponential and weibull. The censoring variable distribution is considered to be same for all blocks and it is taken as exponential.

Modeling and Prediction of Food Crops in Major SAARC Countries

P K Sahu
Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, Nadia, West Bengal, India

Since its formulation in the seventies of the last century, Box- Jenkins methodologies are playing great roles in modeling and forecasting arena, especially to the planners of different countries. To meet the food and nutritional demand of ever-increasing world population in a sustainable manner has remained a great challenge to the planners. Present work is an attempt to study the sustainability in production of major food crops (cereals and vegetables) using measures of sustainability and to model the production behaviour of the same using B-J methodology for major SAARC countries, namely, India, Pakistan, Bangladesh and Nepal. Attempt has also been made to forecast the production potentials of these countries in years to come to supply the food for FAO projected populations of these countries. Taking into consideration the loss of these foods due to various reasons, the study suggests for immediate improvement in per hectare production of these crops.

A Comparative Study on Similarity Search in Time Series

S.Selvakumar∗, K.Sentharamaikannan
Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India

The dimensionality reduction and similarity measures are the basis of time series research, for improving the efficiency and accuracy of time series data mining. In this paper, time series similarity analysis which involves dimensionality reduction, fuzzy c-means and k-means clustering are applied for two different stocks. Mean Absolute Deviation (MAD) is carried out for reducing dimensions and similarity can be found by measuring the distances between time series. The
comparative study will be carried out between k-means and fuzzy c-means clusters are adopted and the efficiency of search and cluster analysis results is carried out graphically and numerically.

A Study on Survival Analysis and Its Exposure to Stochastic Processes

G.Suresh*, K.Senthamaraikannan
Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India

The Survival analysis is one of the vital aspect of the direct application of biostatistical technique in medical practice. Many recent advances has been increased rapidly in the field of medical statistics for the last two decades. Since, the survival analysis is the basic feature of the branch of biostatistics, its significant role implies the incompleteness or censoring of the data which is followed up a patient survival times till his death or because of a patient withdrawing himself from clinical trials and the data becomes incompleteness or censored. The paper discusses the applications of stochastic processes oriented technique related to the significance of life table functions in survival analysis. The methodology for analyzing survival times are discussed with a suitable illustration, using Kaplan-meier method.

Distribution of Products and Ratios of Independently Distributed Pathway Random Variables and Their Properties

Naiju M.Thomas
Centre for Mathematical Sciences, Pala Campus, Kerala, India

A study on the distributions of independently distributed pathway random variables is made and the density functions are obtained in terms of H-functions. G-functions become special cases in the density functions. Mixed products and ratios of independently distributed pathway random variables are also considered. Some of their properties are also considered.
Relative Ranking of Predictors Using a Mixed Model Approach

Soundararajan Govindaraju\textsuperscript{1}, Suresh Veluchamy\textsuperscript{2*}

1. Madras Christian College, Madras, India 2. Einksoft Technologies Private Limited, India

In larges scale surveys, respondents are asked to rank a few listed items. The data is multivariate in nature, with variables representing, for example, various scores given by respondents. Determining the relative importance of these highly correlated variables will lead to better understanding of respondents needs. However, Analysis of Ranking order queries, at times, are more complex, especially when the highly preferred ranking variables fail to discriminate groups. In this paper, we describe a novel, yet simple method to discriminate groups using ranking order preferences. The method determines the relative importance of highly correlated variables after taking multicollinearity problem into account. We illustrate the method using a simulated example in market research.