

International Conference on Recent Advances in Statistics

January 04 - 06, 2005

Abstract



Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Key Note Address:

PROBABILITY MODELS FOR NONNEGATIVE DATA WITH APPLICATIONS IN SURVIVAL ANALYSIS AND RELIABILITY

INGRAM OLKIN
Stanford University, USA

Plenary Talk:

ORDERED RANKED SET SAMPLING AND INFERENCE ISSUES

N. BALAKRISHNAN
McMaster University, Canada

ABSTRACT: Ranked Set Sampling (RSS) and inferential methods based on RSS have been studied quite extensively in the literature. In this talk, I will use the theory of order statistics from i.i.d. variables to introduce the concept of Ordered Ranked Set Sampling (ORSS) and discuss its distributional properties and then use these to develop both parametric and non-parametric inferential methods based on ORSS. I will then illustrate that inference based on ORSS is considerably more efficient than the corresponding inferential procedures based on RSS.

Session 1:

ESTIMATION OF COMPETING RISKS WITH GENERAL MISSING PATTERN IN FAILURE TYPES *

ANUP DEWANJI
Indian Statistical Institute, Calcutta, India

ABSTRACT: In competing risks data, missing failure types (causes) is a very common phenomenon. In this work we consider a general missing pattern in which, if a failure type is not observed, one observes a set of possible types containing the true type along with the failure time. We first consider maximum likelihood estimation with missing at random assumption via the EM algorithm. We then propose a Nelson-Aalen type estimator for situations when certain information on conditional probability of the true type given a set of possible failure types is available from the experimentalists. This is based on a least squares type method

using the relationships between hazards for different types and hazards for different combinations of missing types. We conduct a simulation study to investigate the performance of this method which indicates that bias may be small, even for high proportion of missing data, for sufficiently large number of observations. The estimates are somewhat sensitive to misspecification of the conditional probabilities of the true types when the missing proportion is high. We also consider an example from an animal experiment to illustrate our methodology.

* *This is a joint work with Debasis Sengupta, Indian Statistical Institute, Calcutta.*

A SEMIPARAMETRIC MIXTURE MODEL FOR ANALYZING CLUSTERED COMPETING RISKS DATA

KALYAN DAS

University of Calcutta, India

ABSTRACT: A very general class of multivariate life distributions is considered for analyzing failure time clustered data that are subject to censoring and multiple modes of failure. Conditional on cluster specific quantities, the joint distribution of the failure time and event indicator can be expressed as a mixture of the distribution of time to failure due to a certain type (or specific cause) and the failure type distribution. We assume here, the marginal probabilities of various failure types are logistic functions of some covariates. The cluster specific quantities are subject to some unknown distribution that causes frailty. The unknown frailty distribution is modeled by using nonparametric Dirichlet Process (DP). In such a semiparametric setup, a hybrid method of estimation is proposed based on the i.i.d. Weighted Chinese Restaurant(WCR) algorithm that helps us generate observations from the predictive distribution of the frailty. The Monte Carlo ECM (MCECM) algorithm plays a vital role for obtaining the estimates of the parameters that assess the extent of the effects of the causal factors for failures of a certain type. A simulation study is conducted to study the consistency of our methodology. The proposed methodology is used to analyze a real dataset on HIV infection of a cohort of female prostitutes in Senegal.

Key Words: Frailty model, Monte Carlo ECM algorithm, Weighted Chinese Restaurant Algorithm, Dirichlet Process.

ACCELERATED LIFE TESTING IN THE PRESENCE OF COMPETING *

ISHA DEWAN

Indian Statistical Institute, New Delhi, India

ABSTRACT. In accelerated life testing, the products are tested at high stress conditions and the results are used to draw inferences about the product at the normal stress condition.

However, the system can fail due to more than one cause, for example, fatigue specimens of a certain sintered super alloy can fail from a surface defect or an interior one; in ball bearing assemblies a ball or the race can fail and so on; a semiconductor device can fail at a junction or at a lead and cars can fail due to electrical failure or mechanical failure. Hence, studying accelerated life testing data in the presence of competing risks is essential.

Here, we attempt to combine the competing risks models and accelerated failure time models to propose some likely models for the accelerated life test in the presence of competing risks.

** This is a joint work with Sangita Kulathinal, currently visiting Indian Institute of Technology Kanpur.*

Session 2:

THE SKEW-NORMAL DISTRIBUTION

RAMESHWAR D. GUPTA

University of New Brunswick, Saint John, Canada

ABSTRACT: The univariate skew-normal distribution was introduced by Azzalini (1985, Scand. J. of Stat.) as a natural extension of the classical normal distribution to accommodate asymmetry. In this talk we shall discuss various properties and some problems associated with estimation of the skew parameter. Also, we shall examine various extensions of this model. Finally, we shall show the connection between the class of multivariate skew-normal distributions developed by Azzalini and his co-workers and the class of Liouville distributions studied by Gupta and Richards.

ESTIMATION OF A NORMAL DISPERSION MATRIX USING SHRINKAGE CORRELATION ESTIMATORS

NABENDU PAL

University of Louisiana at Lafayette, USA

ABSTRACT: Estimation of a normal dispersion matrix in a decision theoretic set-up has received attention from many researchers over the past three decades. However, relatively less attention was paid to estimation of correlation coefficients (simple correlation coefficient, intra-class correlation coefficient, coefficient of multiple determination, etc.). In the recent past, we have shown that shrinkage correlation estimators perform better under risk as well as Pitman Nearness criterion. In this paper we have constructed improved dispersion matrix estimators by using shrinkage correlation estimators thereby bridging the gap between improved dispersion matrix estimation and improved correlation estimation.

ESTIMATING RESTRICTED NORMAL MEAN *

SOMESH KUMAR

Indian Institute of Technology Kharagpur, India

Let X_1, X_2, \dots, X_n be a random sample from a normal population with the mean μ and variance σ^2 . In many practical situations, the parameter μ is known a priori to lie in a symmetric interval $[-m, m]$, $m > 0$. The sample mean \bar{X} , which is the usual maximum likelihood estimator (MLE) for estimating μ , is inadmissible in this case. It is due to the fact that it may take values outside the range $[-m, m]$. It is also not minimax with respect to a scale invariant loss function. Casella and Strawderman (Ann. Stat. 9, 1981, pp. 870-878), Bickel (Ann. Stat., 9, 1981, pp. 1301-1309) and Gatsonis, MacGibbon and Strawderman (Stat. Prob. Lett., 6, 1987, pp. 21-30) have suggested various estimators for this problem. They have also studied the minimax value in the revised situation. In this paper, we obtain some new estimators of μ for both the cases when the variance σ^2 may be known and unknown. A study of the comparative risk performance of all these estimators is also taken up.

* *This is a joint work with Yogesh Mani Tripathi, Indian Institute of Technology Kharagpur.*

Session 3:

BAYESIAN ADAPTIVE BIASED-COIN DESIGNS FOR CLINICAL TRIALS WITH NORMAL RESPONSES *

ATANU BISWAS

Indian Statistical Institute, Calcutta, India

ABSTRACT: Adaptive designs are used in phase III clinical trials for skewing the allocation pattern to wards the better treatments. We use optimum design theory to derive a skewed Bayesian biased-coin procedure for sequential designs with continuous responses. The skewed designs are used to provide adaptive designs, the performance of which is studied numerically and theoretically. Important properties are loss and the proportion of allocation to the better treatment.

* *This is a joint work with Anthony C. Atkinson.*

A BAYESIAN APPROACH TO NON-IGNORABLE MISSING RESPONSE IN GENERALIZED LINEAR MIXED MODELS

CHRISTIAN HEUMANN

LMU Muenchen, Germany

ABSTRACT: Missing values are an important special case of coarsened data. Most often, cases containing missing values in some of the studied variables are simply ignored in the

analysis (in practice). As a rule of thumb we can say that whenever the studied outcome or response variable is missing and the probability of missing is dependent on that outcome variable, simple methods like complete case analysis, etc. can lead to a biased estimation. In this talk we treat the problem of non-ignorable missing response in generalized linear mixed models (GLMM), i.e. models with random effects, for non-normal, especially binary data. This type of models is often used for longitudinal or clustered binary data. Recently, a connection of mixed models with smoothing has been formulated and thus the applied methods may be of interest beyond what is studied in the talk. Direct Maximum Likelihood estimation is complicated and therefore a Monte Carlo EM algorithm has been proposed in the literature. While point estimates from such an algorithm are usually reliable when the log-likelihood behaves well, variance estimates are hard to get. Missing values are complicating the estimation process additionally so that a Bayesian approach may be a valuable and attractive alternative. In the talk we develop such an approach and show how modern Markov Chain Monte Carlo methods can be applied. In a number of simulation studies we examined the performance of the so called hybrid Monte Carlo algorithm which avoids the random walk behavior of simple MCMC algorithms by using auxiliary variables and the derivative of the posterior distribution. The mixing behavior is observed to be good which reduces the needed length of the Markov Chain considerably and thus the runtime.

Key Words: Missing data, generalized linear mixed models (GLMM), non-ignorable, Bayesian estimation, hybrid Monte Carlo.

CLUSTERING OF FUNCTIONS BY BAYESIAN WAVELET METHODS

BANI K. MALLICK

Texas A & M University, USA

ABSTRACT: We propose a Bayesian nonparametric model based on the Dirichlet process for clustering functional data. A wavelet basis is used for good localization in time and frequency. The clustering characteristics of the model and the effects of functional properties are studied. We extend the model to handle missing data. Experimental results for synthetic data and a real dataset from Microarray study show that the scope of the application is large.

Session 4:

SIMULTANEOUS CONFIDENCE INTERVALS FOR THE SUCCESSIVE RATIOS OF SCALE PARAMETERS

AMAR NATH GILL

Jaypee University of Information Technology, India

ABSTRACT: We consider k ($k \geq 2$) independent populations or treatments or systems and an absolutely continuous member of location-scale family of distributions, index by the location parameter μ_i ($-\infty < \mu_i < \infty$) and scale parameter θ_i ($\theta_i > 0$), is used to model the

observations from the i -th population, $i = 1, \dots, k$. It is assumed that the data from the i -th population are either contaminated by outliers or doubly censored, $i = 1, \dots, k$ and that there is suspicion that the scale parameters of the populations satisfy the simple ordering $\theta_1 \geq \dots \geq \theta_k$. Motivated by such practical situations, in this paper, we propose one sided and two sided simultaneous confidence intervals for the successive pair-wise ratios $\frac{\theta_i}{\theta_{i+1}}$, $i = 1, \dots, k - 1$, of the scale parameters, with a pre-specified confidence level. A recursive method is derived to compute the critical constants and then applied to obtain the critical constants for two parameter exponential probability models.

Key Words and Phrases: Pair wise comparisons; Critical points; Numerical integration; Simultaneous confidence intervals.

RESTRICTED SUBSET SELECTION FROM SEVERAL NORMAL POPULATIONS WITH A GOAL OF SELECTING AT LEAST ONE OF THE t BEST: KNOWN COMMON VARIANCE CASE

S. PANCHAPAKESAN

Southern Illinois University, USA

ABSTRACT: Consider k normal populations with unknown means and a common known variance. Our goal is to select a non-empty subset of these populations whose size does not exceed a specified bound m ($1 \leq m \leq k - 1$) so that one of the populations associated with the t largest means is included in the selected subset with a guaranteed probability. Our formulation depends on whether $m \leq k - t$ or not. In each case, we discuss two different procedures.

INESTIMABILITY OF ASSOCIATION PARAMETER BASED ON A BROKEN SAMPLE

ZHIDONG BAI

North East Normal University, PRC

and

National University of Singapore, Singapore

ABSTRACT: Suppose that (X_i, Y_i) , $i = 1, 2, \dots, n$ are *i.i.d* random vectors with uniform marginals and a certain joint distribution F_ρ , where ρ is a parameter. However, the X 's and Y 's are observed separately so that the pairing information is missing. Can ρ be consistently estimated? This is an extension of the problem considered in DeGroot and Goel (1980) which focused on the bivariate normal distribution with correlation ρ . In this study, we have proved that there are no consistent estimators of ρ under quite general conditions. When these conditions do not hold, we also give some examples that consistent estimators exist.

Session 5:

APPLICATIONS OF STATISTICS IN FINANCE

SANKARSHAN BASU

Indian Institute of Management, Bangalore, India

ABSTRACT: Statistics have been used to solve various problems in financial sphere for a long time. In fact the first real documented application of statistics used in some form of financial problem dates back to the early 1700's when fairly advanced statistical tools and probability theory were used to lay the foundation of actuarial and insurance mathematics. Since then, the use of statistics in problems in finance have been growing steadily. However, a real documented application of statistics in finance has happened for the last three decades.

This paper looks at two broad aspects:

1. First of all, it studies the various different areas in which statistics and statistical methods have been used to solve various financial problem (both at a theoretical level as well at a practical level). These applications include option pricing (Black and Scholes (1973), Hull and White (1987), Rogers and Shi (1995) and Basu (1999) among others), interest rate modeling (Vasicek (1979), Cox, Ingersoll and Ross (1985), Brace, Gatarek and Musiela (1997) and Basu (1999) among others) as well as insurance related issues (Cox (1955), Dassios (1987), Jang (1998), Basu and Das (2003) and Basu, Dutta, Gupta and Bhatnagar (2004) among others).
2. The second part of this paper is devoted to some discussion of the various possible applications of statistical techniques in various other aspects of finance as well as some open ended problems. This part is more directed at looking at possible research topics in finance where there is a significant application of advanced statistical techniques.

REFERENCES:

1. Basu, S. (1999); Approximating functions of integrals of Log- Gaussian processes: Applications in Finance; Ph.D. Thesis, London School of Economics, University of London.
2. Das, S. and Basu, S. (2003); "A Markov Chain Based No - Claim Discount System"; *Journal of Insurance and Risk Management*, Volume II, Issue 3, 61 - 72.
3. Basu, S., Dutta, G., Bhatnagar, A. and Gupta, R. (2004); "Usage of Mathematical Models and Operations Research Techniques in the Insurance Industry" *Journal of Insurance and Risk Management* forthcoming in the December 2004 issue.

4. Black, F. and Scholes, M. (1973); “The pricing of options and corporate liabilities”; *Journal of Political Economy*, 81, 637 - 659.
5. Brace, A., Gatarek, D. and Musiela, M. (1997); “The market model of interest rate dynamics”; *Mathematical Finance*, 7(1), 127 - 154.
6. Cox, D.R. (1955); “Some statistical methods connected with series of events”; *Journal of the Royal Statistical Society, Series B*, 17, 129 - 164.
7. Cox, J. C., Ingersoll (Jr.), J. E. and Ross, S. A. (1985); “A theory of term structure of interest rates”; *Econometrica*, 53(2), 385 - 407.
8. Dassios, A. (1987); “Insurance, storage and point processes : An approach via piecewise deterministic Markov processes”; Ph.D. thesis, Imperial College, University of London.
9. Hull, J. and White, A. (1987); “ The pricing of options on assets with stochastic volatilities”; *The Journal of Finance*, XLII 2, 281 - 300.
10. Jang, J. (1998); “Doubly stochastic point processes in reinsurance and pricing of catastrophe insurance derivatives”; Ph.D. thesis, London School of Economics, University of London.

RELATIONSHIP BETWEEN MONEY, OUTPUT AND PRICE: REVISITED USING WAVELET FILTERING *

SHARMISHTHA MITRA
Reserve Bank of India, India

ABSTRACT: In this paper, we study the econometric relationship between money, output and price for the Indian economy. We explore their dynamic causal relationship under the Granger-Causality framework and interactions between these primary inputs for central bank policy-making in a structural vector autoregression framework. The much-studied relationship between the three primary indicators of the economy is studied with the help of the recently developed wavelet multiresolution filtering technique. Rather than studying their econometric relationships at the original series level, as is usually done, we first decompose the observed series using wavelet decomposition technique at various scales of resolution and study relationships among components of the decomposed series matched to its scale. Such an analysis reveals interesting aspects of the interrelationships among the three fundamental economic variables.

* *This is a joint work with Amit Mitra, Indian Institute of Technology Mumbai.*

SOME ROBUST STATISTICAL METHODS FOR FINANCE: A REVIEW AND APPLICATIONS

PREM P. TALWAR

University of Alberta, Canada

ABSTRACT: The statistical methods are very widely used in finance to study important issues such as the small firm effect (Roll 1981), Risk Premia (Ibbotson and Sinquefeld 1976), estimates of security risk (Bollerslev 1987 and Talwar 1996), and the testing of the two parameter capital asset pricing model (Tinic and West 1986). It is well known that security price data and commodity prices do not follow the normal distribution as models of randomness in the data. Mandelbrot (1963) suggested that the security returns are generated by a member of the symmetric stable family of distributions. Blattberg and Gonedes (1974) have suggested the use of the student t distribution as a model for stock returns. Akgiray and Booth (1988) find that the distributions of returns are skewed and leptokurtic. Hinich and Patterson (1985) observe that daily stock returns are generated by a nonlinear, non-Gaussian process.

In this presentation, I will review selected robust statistical methods, Talwar (1981, 1983, 1996), Hsu (1979) and Pettitt (1979), and discuss their applications to some recent financial data.

Session 6:

COMPUTATION OF THE K-TH NEAREST NEIGHBOR ESTIMATE OF MOLECULAR ENTROPY USING PARALLEL PROCESSING *

E. JAMES HARNER

West Virginia University, USA

ABSTRACT: Entropy is a statistical measure of the random fluctuations in molecules and its estimation is important for investigating the stability of molecular conformations, for modeling the binding of ligands to proteins, and for studying issues relating to drug designs. Singh et al. (American Journal of Mathematical and Management Sciences, 2003, 23, 301-321) introduced a nonparametric approach for estimating entropy using the kth nearest neighbor distances among sample points, which extends the first nearest neighbor approach of Kozachenko and Leonenko (Problems of Information Transmission, 1987, 23, 95- 101). The entropy of a molecule depends on random fluctuations in its internal coordinates. The high dimensionality of the internal coordinates and a large number of observations on these coordinates cause computational challenges for computing the kth nearest neighbor estimate of entropy of large molecules. We are experimenting with computing the kth nearest neighbor distances for the sample points on a high-performance computer having a large number of parallel processors (nodes). On each processor, we use the Approximate Nearest Neighbor

(ANN) method (Arya et al.(1994), Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, 573-582) for computing kth nearest neighbor distances. ANN builds a tree structure and searches it for nearest neighbors. In order to search for nearest neighbors concurrently using high-performance computing with many processors, we have developed a program to duplicate the search trees on multiple processors (slave nodes) and each node computes the kth nearest neighbor distances for part of the data. When the dimensionality of the coordinates and the number of observations becomes large, the gain in performance becomes obvious. Using the ANN method, the whole data set needs to be loaded into the memory to build the tree. Therefore, the maximum size of the data set that ANN can handle is limited by the size of available memory. We also developed a program that will break the data set into parts and build tree structures on the slave nodes. To find the kth nearest neighbor of a point, the program will search all the trees concurrently and then merge all the results to get the kth nearest neighbor for the data point. The program will enhance the capability of handling extremely large data sets, such as those obtained using molecular dynamics simulation, and it will improve the efficiency of computing the kth nearest neighbor distances for the sample points.

* *This is a joint work with Jun Tan, Shengqiao Li and Harshinder Singh West Virginia University, Morgantown.*

NEAREST NEIGHBOR ESTIMATES OF ENTROPY OF RANDOM FLUCTUATIONS IN MOLECULES

HARSHINDER SINGH

West Virginia University, USA

and

National Institute for Occupational Safety and Health, Morgantown, USA

ABSTRACT: Estimation of entropy of random fluctuations in the internal coordinates (bond lengths, bond angles and dihedral angles) of a molecule is an important problem in molecular sciences for studying their structures, properties and functions. Large molecules have a large dimension of internal coordinates and the entropy estimation for such molecules is a challenging problem. The traditional approach is based on assuming a multivariate normal distribution for the coordinates of the molecule. However, the assumption of normality does not hold in general and particularly it does not hold at higher temperatures. We discuss a nonparametric approach to the estimation of entropy which is based on k-th nearest neighbor Euclidean distances between the sampled observations, where k is positive integer less than the total number of observations (Kozachenko and Leonenko. (1987), Problems of Information Transmission, Singh et al (2003), American Journal of Mathematical and Management Science). Bond lengths and bond angles are rather hard coordinates and therefore a significant contribution to entropy comes from the random fluctuations in the dihedral angles of a molecule. In many practical applications, observations only on the dihedral angles are considered for estimation of entropy of the molecule (Demchuk and Singh (2001), Molecular

Physics). Thus, we also discuss k -th nearest neighbor estimates of entropy for multivariate circular distributions where we use circular distances in place of the Euclidean distance (Misra et al.; 2004, Preprint).

SOME RESULTS ON GENERALIZED RESIDUAL ENTROPY *

ASOK K. NANDA

Indian Institute of Technology Kharagpur, India

ABSTRACT: Shannon's entropy plays an important role in the context of information theory. Since this entropy is not applicable to a system which has survived for some units of time, the concept of residual entropy has been developed in the literature. Here we generalize the residual entropy by choosing a convex function with $(1) = 0$. In this paper, some orderings and aging properties have been defined in terms of the generalized residual entropy function and their properties have been studied. Quite a few results available in the literature have been generalized and some distributions (viz. uniform, exponential, Pareto and power series) have been characterized through the generalized residual entropy.

* *This is a joint work with Prasanta Paul, Indian Institute of Technology Kharagpur.*

Session 7:

MONTE CARLO VERSION OF THE EM ALGORITHM

T. KRISHNAN

Systat Software Asia Pacific Limited, India

ABSTRACT The Expectation-Maximization (EM) algorithm is an iterative method of computing maximum likelihood estimates (MLE) and maximum a posteriori (MAP) estimates in problems where the model and/or the data make Newton-Raphson-like methods messy. Application of the algorithm involves formulation of an associated "nice" problem with simpler likelihood and hence simpler MLE. Each iteration involves two steps—the Expectation step (E-step) and the Maximization step (M-step). But in some problems the E-step of deriving the expected value of the log-likelihood of the "nice" problem may be analytically intractable. In other problems, the maximization of even the simpler problem may be difficult. Either or both of these difficulties can often be overcome by resorting to Monte Carlo procedures. Such an approach gives rise to a variety of Monte Carlo versions of the EM algorithm, including some Markov chain Monte Carlo (MCMC) algorithms. In this presentation, we review these algorithms.

STATISTICAL LEARNING FROM DNA WORDS

PROBAL CHAUDHURI

Indian Statistical Institute, Calcutta, India

ABSTRACT: I will start with some examples from evolutionary biology to demonstrate how distributions of various words in DNA sequences can be used to learn important facts in evolutionary biology. The statistical analysis used involves an interesting variable selection problem. This motivated developing certain probabilistic models for replication of DNA sequences which will be discussed.

ROLE OF ANN AND STATISTICAL TECHNIQUES IN THE CLASSIFICATION OF BENIGN AND MALIGNANT TUMORS *

KULDEEP KUMAR

Bond University, Australia

ABSTRACT: Digital mammography is one of the most suitable methods for early detection of breast cancer. It uses digital mammograms to find suspicious areas containing benign and malignant microcalcifications. However, it is very difficult to distinguish benign and malignant microcalcifications. This is reflected in the high percentage of unnecessary biopsies that are performed and many deaths caused by late detection or misdiagnosis. A computer based feature selection and classification system can provide a second opinion to the radiologists in assessment of microcalcifications. The research in this paper proposes and investigates a neural-genetic algorithm for feature selection in conjunction with neural and statistical classifiers to classify microcalcification patterns in digital mammograms. The obtained results show that the proposed approach is able to find an appropriate feature subset and neural classifier achieves better results than two statistical models.

* *This is a joint work with Ping Zhang, Bond University, Australia.*

FREQUENCY ESTIMATION OF SIGNAL PROCESSING MODELS USING GENETIC ALGORITHM *

AMIT MITRA

Indian Institute of Technology Mumbai, India

ABSTRACT: We consider the problem of frequency estimation of two important signal processing models, the complex valued quadrature signals model and the sinusoidal signals model. We propose genetic algorithm and outlier-insensitive criterion function based iterative techniques for the frequency estimation problem. The proposed methods significantly do not depend on the initial guess values otherwise required for the different iterative frequency estimation methods. A two stage computationally efficient algorithm is also proposed. In

the simulation studies and real life data analysis, it is observed that the proposed genetic algorithm based estimators are able to resolve frequencies of the sinusoidal model with high degree of accuracy. It is also observed that the methods have fairly high breakdown point with respect to different types of outliers present in the data.

* *This is a joint work with Debasis Kundu, Indian Institute of Technology Kanpur.*

Session 8:

UNIFORMITY IN FACTORIAL DESIGNS WITH MIXED LEVELS *

KASHINATH CHATTERJEE
Visva-Bharati University, Santiniketan, India

ABSTRACT: The uniformity can be utilized as a measure for comparing factorial designs. Fang and Mukerjee (2000) and Fang et al. (2002) found links among the uniformity in terms of centered L_2 -discrepancy or wrap-around L_2 -discrepancy, orthogonality and aberration for regular symmetric factorials. We in this paper extend their results to asymmetric factorials. Furthermore, some lower bounds under different non-uniformity measures are obtained for asymmetric factorials and two new ways of construction of factorial designs with mixed levels are proposed.

* *This is joint work with Kai-Tai Fang, Hong Kong Baptist University, Hong Kong, and Hong Qin, Hong Kong Baptist University, Hong Kong.*

A NOTE ON THE EFFICIENCY OF SOME DESIGNS FOR THE PARTITION PROBLEM

TUMULESH KUMAR S SOLANKY
University of New Orleans, USA

ABSTRACT: We consider the problem to partition a set of normal populations with respect to a control population. We would construct some efficient designs for the problem and also study the robustness of proposed and some existing procedures under small departures from assumptions.

APPLICATION OF SPATIAL STATISTICAL MODELS IN AGRICULTURAL RESEARCH AND MANAGEMENT *

M. NARAYANA REDDY
Natl. Acad. of Agri. Res. Mgmt, Hyderabad, India

ABSTRACT: Spatial statistical modeling is one of the priority areas in agricultural research. Advances that have taken place in spatially referenced database management systems such

as GIS (Geographical Information Systems), spatial analysis has become an active application in agricultural research and management to provide better insight into the spatial phenomenon from field to regional level. Spatial statistics in particular geo-statistical models and spatial interaction models with GIS are valuable tools to characterize spatial and temporal patterns of agricultural practices. These patterns aid in taking better decisions on sustainable agricultural practices. Precision agriculture is a fundamental paradigm shift in agricultural management. In general soil nutrient variability maps which are basic input to precision agriculture produced in GIS using spatial analysis ignores spatial dependence resulting in under sampling and unwarrantly optimistic confidence intervals in parameter estimation and prediction. Geo-statistical models applied to spatial databases on various soil and climatic spatial variables produce scientifically valid input maps required for precision agriculture. As spatial data in GIS invariably contain random errors, geostatistical tools model these random errors to arrive at scientific inferences. In this paper the goodness of fit of geo-statistical techniques to produce scientifically valid maps is examined with yield data collected from uniformity trial on four crops. Spatial models also proved to be efficient in the analysis of field experiments. Efficiency of spatial statistical analysis is studied in relation to the traditional analysis assuming independence among errors using the data from uniformity trials. The application of spatial interaction models to study the plant competition is also demonstrated.

* *This is a joint work with G.K. Shukla, Indian Institute of Management, Lucknow.*

MULTIVARIATE SPATIAL MODELS FOR SMALL AREA ESTIMATION *

BHARAT BHUSAN SINGH

Central Statistical Organization, New Delhi, India

ABSTRACT: Small Area Estimation techniques play an important role in survey sampling due to growing demands for reliable small area estimates from planners, researchers and decision makers. The estimates of the small area parameters can preferably borrow strength from the data on related multiple characteristics and/or auxiliary variables from other neighboring areas through appropriate models. Fay (1987) and Datta, Day and Basava (1999) proposed multivariate regression for small area estimation of multiple characteristics.

Small area characteristics have the spatial dependence in terms of neighborhood similarities. Estimates based on spatial regression model have been found to be more precise and efficient than the estimates under linear mixed models. Here the strength is borrowed from the similar small areas through two common parameters viz. regression parameter β and autocorrelation parameter ρ . To estimate small area mean vectors of multiple characteristics, multivariate variance components models have been proposed in literature. The degree of correlation coefficients between related characteristics, under small area model framework, improves the small area estimates for each of the characteristics. This aspect of multivariate characteristic has been exploited in this paper, in the form of a multivariate spatial model.

Here a combination of correlation coefficients between the sampling design errors and between the small area random effects, along with different autocorrelation parameters in a multivariate setup, have been used to improve the small area estimates of each of the characteristics. We use this approach to find empirical best linear unbiased predictor (EBLUP) and the associated mean squared and product error (MSPE) of prediction. We have also found out the second order approximation to the estimate of the MSPE. Data on monthly per capita expenditure for food and non-food items, as obtained through a large sample survey carried out by the National sample Survey Organization (NSSO), Ministry of statistics and Program Implementation, Government of India, have been used to validate the models.

* *This is a joint work with G.K. Shukla, Indian Institute of Management, Lucknow.*

Session 9:

INVARIANT ESTIMATOR IN A QUADRATIC MEASUREMENT ERROR MODEL *

ALEXANDER KUKUSH

Kiev National Taras Shevchenko University, Ukraine

ABSTRACT: An adjusted least squares (ALS) estimator is derived that yields a similarity-invariant and consistent estimate of the parameters of an implicit quadratic measurement error model (IQMEM). Consistency means that the estimate converges to the true value of the parameter, as the sample size tends to infinity. In addition, a consistent estimator for the measurement error noise variance is proposed. Important assumption are: (1) all errors are i.i.d. and (2) the error distribution is rotation-invariant. The estimators for the quadratic measurement error model are used to estimate consistently conic sections and ellipsoids. In the IQMEM, the ordinary least squares (OLS) estimator is inconsistent, and due to the non-linearity of the model, the orthogonal regression (OR) estimator is inconsistent as well. Simulation examples, comparing the ALS estimator with the OLS method and the OR method, are discussed for the ellipsoid fitting problem. The results are joint with The consistency is shown in [1], and the numerical algorithm is proposed in [2].

REFERENCES:

- [1] A.Kukush, I.Markovsky, and S.Van Huffel, Consistent estimation in an implicit quadratic measurement error model. Computational Statistics and Data Analysis, 2004, 47, 1, 123-147.
- [2] I.Markovsky, A.Kukush, and S.Van Huffel, Consistent least squares fitting of ellipsoids. Numerische Mathematik, 2004, 98, 1, 177-194.

* *This is a joint work with S.Van Huffel, I.Markovsky (Belgium) and S.Shklyar (Ukraine).*

EFFICIENCY COMPARISON OF SEVERAL ESTIMATORS OF A MEASUREMENT ERROR MODEL

HANS SCHNEEWEISS
Munich University, Germany

ABSTRACT: The estimators to be compared are the naive estimator, which is biased, and the consistent estimators CS (corrected scores) and QS (quasi scores) as well as ML (maximum likelihood) and SS (a simplified version of QS). The asymptotic covariance matrices are compared to each other, in particular for small measurement error variances, where the consistent estimators all become almost equally efficient, and in some cases (polynomial and Poisson model) for large measurement errors, where QS seems to be the winner. QS, however, is sensitive to mis-specifications of the regressor distribution. The bias resulting from these mis-specifications is also investigated.

COMPARISON OF SEQUENTIAL EXPERIMENTS FOR ESTIMATING THE NUMBER OF CLASSES IN A POPULATION

TAPAN K. NAYAK
George Washington University, USA

ABSTRACT: This talk will deal with stopping rules in sequential sampling from a population with an unknown number of classes or species. Adopting Blackwell's (1951) ideas we define "more informative stopping rule," which induces a partial ordering of all stopping rules. Some consequences of more informativeness, and certain complete class results will be discussed. Necessary and sufficient conditions for a stopping rule to yield complete sufficient statistics will be presented and explained geometrically. Unbiased estimation of the class size and some functions of it will also be discussed.

Session 10:

A TEST FOR SEPARATE FAMILIES OF HYPOTHESES BASED ON PREDICTIVE SAMPLE REUSE TECHNIQUES *

GIRJA KANT SHUKLA
Indian Institute of Management Lucknow, India

ABSTRACT: Model selection is an important problem in data analysis, especially now with more powerful computational tools and increasing interest in data-mining. Very often models are used to predict future observations. Predictive loss or predicting density approach has been widely used in model choice when the models under consideration are non-nested, but not so much in hypotheses testing, apart from a recent proposal, by Presnell and Boos

(2004). Hypotheses testing can be a cost-effective approach if one model is already in use. We propose a test based on predictive approach using cross-validation to test the departures from one model in the direction of another when the models are non-nested or, in other words, they belong to separate families of hypotheses. Asymptotic theory of the test is developed. Performance of the test for various sample sizes is illustrated with an application. The methodology has important applications as the test can be easily adapted for non-nested regression and non-nested Generalized Linear Models (Bawa, 1994). Parametric bootstrap testing of hypothesis has also been discussed for small sample size. Examples of analyzing real life data are also given.

Keywords: Cross-validation; Model selection; Non-nested models; Separate families of hypotheses; Predicting density.

* *This is a joint work with Nisha Bawa Mahajan Smith Hanley Consulting Group.*

STATISTICS: ITS APPLICATIONS IN DRUG DEVELOPMENT

SHUBHA RANI

B.V. Patel PERD Centre, India

ABSTRACT: Drug research is a unique multi-disciplinary process heading to wards the development of novel therapeutic agents in areas of currently unmet medical need. Although the importance for statistical inputs in the pharmaceutical sciences has always been evident, 35 to 45 years ago its absence posed no significant problems. There was no norm to review possible poor design or possible lack of optimality. The ultimate judge of product development was the consumer. The development of drugs is eventually dependent on their effectiveness and safety in living materials, particularly in humans. High variability of human responses makes the use of statistical inputs a necessity when assessing the drug efficacy and safety in the present scientific environment. It has become essential to enhance the quality of any experiment or study by including statistical inputs at the stage of designing the protocol itself. Statistical design and analysis are essential tools for the pharmaceutical industries to use in order to properly develop drugs that will be judged approvable by regulatory agencies. In many countries, the regulatory authorities for drugs have a well established statistical group which has made its voice heard loud clear. The presentation will highlight the applications of statistics at various stages of drug development. Specifically, the evolution of statistical concept in bioequivalence studies will be discussed.

A STYLOMETRIC ANALYSIS OF KING ALFRED'S LITERARY WORKS *

PARAMJIT GILL

Okanagan University College, Canada

ABSTRACT: After King Alfred the Great (848-899) defeated the Vikings at the battle of Edington in 878, he turned to strengthening his English kingdom of Wessex that had suffered so greatly under the Viking invasions. Most famous was his program for educational

reform. Alfred depicted himself as a philosopher-king, taking up scholarship and making English translations of Latin patristic texts to serve as the basis of education in the English language. For centuries, Alfred the Great was judged to have translated several texts. Many scholars, however, have expressed doubt whether Alfred could have done all of this work. With the availability of the Old English Corpus in electronic form, it is feasible to subject the texts to statistical stylometric analysis. We approach the problem from a Bayesian perspective where key words are identified and frequencies of the key words are tabulated for seven relevant texts. The question of authorship falls into the general statistical problem of classification where several simple innovations to classical agglomerative procedures are introduced. Our results suggest that three translations, "Pastoral Care", "The Consolation of Philosophy" and "The Soliloquies" that have been traditionally attributed to Alfred indeed cluster according frequency of the key words. However, one translation that has been traditionally attributed to Alfred "The First Fifty Prose Psalms" tends to distinguish itself from texts that are known to be Alfredian.

* *This is a joint work with Tim Swartz, Okanagan University College.*

Session 11:

INFERENCE FOR THE SIMPLE STEP-STRESS MODEL *

NANDINI KANNAN

The University of Texas at San Antonio, USA

ABSTRACT:In several applications in Survival Analysis and Reliability, the experimenter is often interested in the effects of extreme or varying stress levels (temperature, voltage, load, pollution levels) on the lifetimes of the experimental units. Step Stress Tests allow the experimenter to increase the stress levels at prespecified times during the experiment. The most common model used to analyze these experiments is the "cumulative damage" or "cumulative exposure" model.

In this talk, we consider a simple step-stress model with only two stress levels. We assume the failure times at the two different stress levels follow exponential distributions. When the stress level changes, the distribution is assumed to be linear exponential which allows us to model a "lag" effect of the stress.

We obtain the Maximum Likelihood estimators of the unknown parameters and propose different techniques for constructing confidence intervals.

Key Words and Phrases: Accelerated Testing; step-stress models; cumulative exposure model; maximum likelihood estimation; bootstrap method; Type-II censoring; spacings; exponential distribution.

* *This is a joint work with N. Balakrishnan, McMaster University, D. Kundu, Indian Institute of Technology Kanpur and H. K. T. Ng, Southern Methodist University*

A BIVARIATE DISCRETE RELIABILITY MODEL *

HARE KRISHNA

Chaudhary Charan Singh University, India

ABSTRACT: In this paper a bivariate geometric distribution is suggested as a lifetime model. Some examples of this model in real life situations are given. Expressions and numerical values for the reliability characteristics of this distribution are derived. The maximum likelihood estimation of parameters and reliability has also been attempted.

* *This is a joint work with Pramendra Singh Pundir, Chaudhary Charan Singh University.*

TOLERANCE INTERVALS FOR GENERALIZED RAYLEIGH DISTRIBUTION *

R. R. KUMBHAR

Kisan Veer Mahavidyalaya, India

ABSTRACT: Aminizadhe (1991) has reported tolerance intervals for Rayleigh distribution based on pivotal quantity. Recently Kundu and Raqab (2003) have discussed different methods of estimations for the parameters of the generalized Rayleigh distribution. In this article we provide an asymptotic upper β -expectation and β -content γ -level tolerance intervals for this distribution. Expected coverage of the proposed tolerance interval is obtained. The performance of the proposed tolerance intervals is assessed through simulation study.

* *This is a joint work with D. T. Shirke, Shivaji University.*

Session 12:

MULTIVARIATE CONTROL CHARTS FOR AUTOCORRELATED DATA*

ANOOP CHATURVEDI

University of Allahabad, India

ABSTRACT: For monitoring an industrial process, various multivariate quality control charts have been proposed in literature with the aim of controlling several correlated variables simultaneously. The main advantage of multivariate control charts is that these charts take advantage of relationship between different variables. While developing these multivariate control charts, the observation vectors on quality characteristics to be monitored are usually assumed as independently distributed. In univariate control charts, to model the correlation structure of different observations, time series models like AR model, MA model or

mixed ARMA model have been widely used and control charts for autocorrelated data have been proposed. The present paper considers the multivariate control charts, when observation vectors on quality characteristics under consideration follow auto-regressive process. Hotelling's multivariate control chart, multivariate CUSUM chart, and multivariate Exponentially weighted moving average (MEWMA) control charts have been discussed. The results have been extended when multivariate autocorrelated observations are available on a two-dimensional grid.

* *This is a joint work with Chandra Gulati, University of Wollongong, Australia.*

PROCESS CAPABILITY ANALYSIS FOR NON-NORMAL DISTRIBUTIONS

D. T. SHIRKE
Shivaji University, India

ABSTRACT: Process capability index is an important measure to evaluate the capability of the manufacturing process. Many of the process capability indices available in the literature are based on assumption of normality of the process output characteristic. Somerville and Montgomery (1996) among others have reported that using normal parametric indices on non-normal process data give unreliable results. Nonparametric process capability indices have been reported by Polansky(1998) and others. In the present talk several approaches to deal with non-normal situation will be discussed. A new concept of Process Capability Function to describe the capability of the process is introduced. Results based on simulation study to investigate performance of the proposed function will be presented.

BONFERRONI CURVE AND RELATED STATISTICAL INFERENCE *

KANCHAN JAIN
Panjab University, India

ABSTRACT: Two measures of income inequality, viz. the Bonferroni Curve and the Bonferroni index are studied. It is shown that they have some relationship with Lorenz Curve and Gini Ratio and certain concepts used in Reliability and Life testing. The variance-covariance structure of the vector of Bonferroni Curve ordinates is obtained and the problem of testing for the Bonferroni Curve and the Bonferroni index is discussed. The exact sampling distribution of the Bonferroni index is derived.

* *This is a joint work with Sangeeta Arora, Panjab University.*

Session 13:

A NATURAL GOODNESS-OF-FIT TEST FOR THE EXPONENTIAL GROWTH CURVE MODEL *

AYANENDRANATH BASU

Indian Statistical Institute, Calcutta, India

ABSTRACT: A natural goodness-of-fit test, based on the property of finite differences, is proposed for the exponential growth curve model (EGCM), which has wide applications in various fields such as biology, ecology, demography, population dynamics, finance, econometrics etc. In the spirit of the Potthoff-Roy model (1964, *Biometrika*) our procedure assumes multivariate normality for longitudinal data at different time points with a one step Markov structure in the correlation matrix. Hill (1968, *Biometrics*) proposed a test for the polynomial growth curve model (PGCM). However, the PGCM does not have any biologically interpretable parameters and its practical use is limited. In comparison, the exponential model is more realistic and has a biological interpretation which might explain the growth process quite well; thus a goodness-of-fit test for the EGCM has substantial practical value. Remarkably, there has so far been no attempt to generalize and extend Hill's approach to the case of simple nonlinear models such as the EGCM, a gap which we hope to partially fill with the present work. The proposed test will aid researchers in several disciplines to reach at a decision about the hypothesis of an EGCM. The performance of the theory developed is illustrated with several sets of real data and through simulation.

* *This is a joint work with Sabyasachi Bhattacharya and Subhadip Bandyopadhyay.*

BAYESIAN INFERENCE FROM CASE-COHORT DATA WITH MULTIPLE END-POINTS *

SANGITA KULATHINAL

KTL - National Public Health Institute, Finland

ABSTRACT: In a case-cohort design a random sample from the study cohort, which is referred to as a sub-cohort, as well as all the cases outside the sub-cohort are selected for collecting extra covariate data. The union of the selected sub-cohort and all cases is referred to as the case-cohort set. Such a design is generally employed when the collection of information on an extra covariate for the study cohort is expensive. An advantage of the case-cohort design over more traditional case-control and the nested case-control designs is that it provides a set of controls which can be used for multiple end-points, in which case there is information on some covariates and event follow-up for the whole study cohort. Here we propose a likelihood approach to analyze such a case-cohort design as a cohort design with missing data on the extra covariate. We construct likelihood expressions when multiple end-points are of interest simultaneously and propose a Bayesian data augmentation method to estimate the model

parameters. A simulation study is carried out to illustrate the method and the results are compared with the complete cohort analysis.

* *This is a joint work with Elja Arjas, University of Helsinki.*

Session 14:

LOOKING FOR A MORE ACCURATE ESTIMATE OF POPULATION ATTRIBUTABLE RISK

SARABASHI BASU

Indian Statistical Institute, Calcutta, India

ABSTRACT: The population attributable risk (PAR) of a disease due to a risk factor quantifies the proportion of the preventable disease burden when the risk factor is removed from the population. PAR was originally proposed for a binary risk factor and a binary disease outcome. Later on logistic regression formulation of PAR was developed to accommodate multi-factorial or ordinal risks. Maximum likelihood estimate (MLE) of PAR and its asymptotic standard error (ASE) are in use for several decades. In this talk we will first review some of the existing methods for estimation of PAR and then indicate how more accurate estimates can be found.

Since PAR is a non-linear function of multinomial probability, estimate of ASE is not precise for Small sample size. Application of re-sampling is advocated to obtain better variance estimates and confidence intervals for both model-based and non model-based MLEs of PAR. The model-based method is extended to take into account ordinality in the disease outcome. Cumulative logit models under proportional odds assumption are proposed and a sequence of threshold-specific PARs are formulated. MLEs of model-based PARs are shown to perform better than non-model based estimates.

MOMENTS OF LORENZ CURVE AND LINEAR MEASURES OF INEQUALITY

SANGEETA ARORA

Dept. of Statistics

Panjab University Chandigarh-160014

ABSTRACT: The relationship of moments of Lorenz Curve with two other families of Inequality measures is established. It is shown that the moments of the Lorenz curve and these two families of inequality measures can be taken as particular cases of general class of linear measures of income inequality with suitable score functions. Each score function defines a particular linear inequality measure and we study the basic properties of such measures using some restriction on score functions. These linear measures of income inequality

can also be put as linear functions of order statistics and asymptotic distribution of some of such measures is derived. Tests of significance for one-sample and two samples based on these inequality indices are suggested. Finally, the simulation work is carried out to find the power of these tests.

Session 15:

MUTUAL INFORMATION ESTIMATES AND ISSUES WITH STANDARD ERROR

APRATIM GUHA

University of California, Berkeley, USA

ABSTRACT: Originally used in the field of information theory, mutual information is increasingly becoming a popular choice among researchers from other fields to measure the strength of association between of two (or more) random processes. However, not much work has been done to assess the standard errors of the estimates of mutual information, Fernandes (2001) is perhaps the only known source, dealing with the asymptotic distribution of the mutual information estimates for two independent continuous time series. We discuss the asymptotic distribution of the mutual information estimates for hybrid processes, e.g. bivariate time series with a continuous and a binary component, and We demonstrate by simulated examples that convergence to the large sample distribution can be very slow in the application of asymptotic standard errors, even on estimates based on moderately large data sets. As an effective alternative, we suggest standard error estimates calculated using blocked bootstrap methods. Simulated examples to support our claim are shown. Real life examples are provided.

ANALYSIS OF TYPE-II PROGRESSIVELY HYBRID CENSORED DATA *

AVIJIT JOARDER

Reserve Bank of India, India

ABSTRACT: The mixture of Type-I and Type-II censoring schemes, called the hybrid censoring scheme is quite common in life-testing or reliability experiments. Recently Type-II progressive censoring scheme becomes quite popular for analyzing highly reliable data. One drawback of the Type-II progressive censoring scheme is that the length of the experiment can be quite large. In this paper we introduce a Type-II progressively hybrid censoring scheme, where the experiment terminates at a pre-specified time. For this censoring scheme, we analyze the data under the assumptions that the lifetime distributions of the different items are independent and exponentially distributed random variables. We obtain the maximum

likelihood estimator of the unknown parameter in an exact form. Asymptotic confidence interval and two bootstrap confidence intervals are also proposed. Bayes estimate and credible interval of the unknown parameter are obtained under the assumption of gamma prior of the unknown parameter. Different methods have been compared using Monte Carlo simulations. One real data set has been analyzed for illustrative purposes.

** This is a joint work with Debasis Kundu, Indian Institute of Technology Kanpur.*

DISCRIMINATING BETWEEN THE WEIBULL AND LOG-NORMAL DISTRIBUTIONS *

ANUBHAV MANGLICK
University of Passau, Germany

ABSTRACT: Log-Normal and Weibull distributions are the most popular distributions for modeling skewed data. In this paper, we consider the ratio of the maximized likelihood in choosing between the two distributions. The asymptotic distribution of the logarithm of the maximized likelihood ratio has been obtained. It is observed that the asymptotic distribution is independent of the unknown parameters. The asymptotic distribution has been used to determine the minimum sample size required to discriminate between two families of distributions for a user specified probability of correct selection. We perform some numerical experiments to observe how the asymptotic methods work for different sample sizes. It is observed that the asymptotic results work quite well even for small samples also. Two real data sets have been analyzed.

** This is a joint work with Debasis Kundu, Indian Institute of Technology Kanpur.*

Session 16:

SOME STRONG LAW RESULTS FOR RANDOM GEOMETRIC GRAPH *

BHUPENDRA GUPTA
Indian Institute of Technology Kanpur, India

ABSTRACT: We study some important properties of exponential random geometric graph and their asymptotic behaviors. We also study the Uniform random geometric and truncated exponential random geometric graph by approximating them appropriate independent spacing models. Also, study the thresholding behavior of uniform exponential random geometric

graphs. We derive exact threshold for the probability of connectivity and some strong law results.

** This is joint work with Srikanth K. Iyer, Indian Institute of Technology Kanpur.*

TESTING PARAMETERS OF ZERO - INFLATED POWER SERIES DISTRIBUTIONS AND APPLICATIONS *

M. K. PATIL

P.V.P. Mahavidyalaya Kavathe, India

ABSTRACT: In this article we consider zero-inflated power series distributions. We provide asymptotic test based on maximum likelihood estimators for testing parameters of the zero-inflated power series distributions. Zero-inflated Poisson, binomial, negative binomial models have been considered as particular cases of the zero - inflated power series distributions and test statistics for each distributions are given. Performance of the proposed test for three distributions is studied using simulation experiments. Applications to real life data have been given. Some further related results are also discussed.

** This is a joint work with D. T. Shirke, Shivaji University.*

EFFECTS OF DIFFERENT CONTROL LEVELS IN EPIDEMIC MODEL FOR NEW EMERGING DISEASES *

ANUJ MUBAYI

Arizona State University, USA

ABSTRACT: In this paper classical epidemiology models are modified to incorporate the effects of control stages (three different quarantine classes). Threshold, equilibria and stability criteria are explored. Model dynamics are also analyzed and the possibility of the existence of periodic solutions via Hopf bifurcation is addressed. Numerical simulations are carried out to verify the theoretical results. Stochastic version of this model is also formulated.

** This is a joint work with Maia Martcheva, University of Florida, Carlos Castillo-Chavez, Arizona State University and Gerardo Chowell, Los Alamos National Lab.*