INDUCTIVE STATISTICS: FOUNDATIONS AND APPLICATIONS

ORGANIZER: LAURA RIFO, UNIVERSITY OF CAMPINAS, BRAZIL

This session presents some of the works produced by the Inductive Statistics CNPq Researh Group, and will be organized in two days.

TALK 1: CATEGORICAL DATA ANALYSIS USING A SKEWED WEIBULL REGRESSION MODEL

SPEAKER: ADRIANO POLPO DE CAMPOS, FEDERAL UNIVERSITY OF SÃO CARLOS, BRAZIL

In this paper, we present a Weibull link (skewed) model for categorical response data arising from binomial as well as multinomial models. We show that, for such types of categorical data, the most commonly used models (logit, probit and complementary log-log) can be obtained as limiting cases. We further compare the proposed model with some other asymmetrical models. The Bayesian as well as frequentist estimation procedures for binomial and multinomial data responses are presented in details. The analysis of two data sets to show the efficiency of the proposed model is performed. This is a joint work with Renault Caron, Debajyoti Sinha, Dipak Dey.

TALK 2: IMPLEMENTATION OF CENSORED REGRESSION MODELS FOR SURVIVAL ANALYSIS

SPEAKER: DANILO L. LOPES, FEDERAL UNIVERSITY OF SÃO CARLOS, BRAZIL

We study the Bayesian censored modeling of several probability distributions that are widely used in reliability and survival analysis. We discuss the implementation a new package for the free software environment R, which provides routines for the Bayesian estimation of several distributions recently introduced in the statistical literature. For some of these distributions, a new parameterization is presented in order to simplify the estimation process. We also perform a simulation study to analyze the quality of the developed procedures. Finally, we discuss on some problems related to the estimation of such complex parametric models.

TALK 3: A NON-INFERIORITY TEST FOR SURVIVAL RESPONSE ASSURING THE CONTROL OF TYPE I ERROR

SPEAKER: JULIANA COBRE, UNIVERSITY OF SÃO PAULO, BRAZIL

The bio-pharmaceutical industry has been developing new treatments that, compared to a standard treatment of proven efficacy, have characteristics such as greater quality of life, lower cost, easier administration, decreased side effects. Even the new treatment being worse in terms of the survival outcome it can be approved if it has such advantages. In this cases a non-inferiority hypothesis test, instead of a superiority or an equivalent test, must be conducted. Often the non-inferiority trials for survival response assume that the data follow a proportional hazard model (PHM). Nevertheless we prove here that if this assumption is not true the usual log-rank based non-inferiority tests inflate the type I error rate, rather when the sample size goes to infinity, the limit of the type I error rate is equal to 1 at the margin of the null space. To circumvent this serious problem, we propose a proportional odds survival model (POSM) based non-inferiority hypothesis test. The main advantage of our test is that it controls the type I error rate when the true model is both PHM and POSM. The sample sizes determination to reach a desired power complete our proposal.

TALK 4: BAYESIAN INFERENCE FOR THE MEMORY IN DIFFUSION PROCESSES

SPEAKER: LAURA RIFO, UNIVERSITY OF CAMPINAS, BRAZIL

We present diverse papers concerning the inference problem for the memory parameter in stochastic processes derived from or related to the Brownian motion (Bm): fractional Bm, fractional Ornstein-Uhlenbeck, Rosenblatt, Gamma modulated processes. These are joint works with Plinio Andrade (University of São Paulo, Brazil), Soledad Torres and Francisco Torres-Avilés (Universidad de Santiago de Chile, Chile), and Ciprian Tudor (Université de Lille, France).

TALK 5: A SIMPLE ALGORITHM FOR MAXIMUM LIKELIHOOD ESTIMATION OF POPULATION SIZE UNDER A HETEROGENEOUS CLOSED CAPTURE-RECAPTURE MODEL

SPEAKER: LUIS ERNESTO BUENO SALASAR, FEDERAL UNIVERSITY OF SÃO CARLOS, BRAZIL

We present a simple EM-based procedure to obtain maximum likelihood estimates of population size under a heterogeneous closed capture-recapture model. Heterogeneity is modeled by assuming a finite discrete mixing distribution for the individual capture probabilities. For this model, it is shown that a constraint on the parameters should be adopted in order to achieve identifiability. Despite the complicated likelihood function, the proposed algorithm can handle easily the maximization without requiring the use of customary

optimization procedures such as Newton-Raphson iterative methods. The proposed algorithm is illustrated on a real data set.

TALK 6: DATA ANALYSIS FOR WEIBULL VARIATES INTERVAL CENSURED

SPEAKER: TERESA CRISTINA MARTINS DIAS, FEDERAL UNIVERSITY OF SÃO CARLOS, BRAZIL

Interval censured data are relatively usual in practice in several areas. For this kind of censure, the occurring times are not exactly known: we only know that the event occurred inside a given time interval (L,R], L<R. In survival analysis, suitable techniques are applied to estimate, by instance, the empirical survival function. Point and interval estimates for the model parameters can be obtained under both Bayesian and frequentist approaches. In this work, we analyze data under interval censure, assuming they are distributed according to a Weibull distribution. Estimates for the related parameters are obtained by an EM algorithm, the classical approach, and a Metropolis-Hastings algorithm, the Bayesian approach, under non-informative prior specification.

TALK 7: CLASSIFYING THE ORIGIN OF ARCHAEOLOGICAL FRAGMENTS WITH BAYESIAN NETWORKS

SPEAKER: VICTOR FOSSALUZA, UNIVERSITY OF SÃO PAULO, BRAZIL

Classification of archeological fragments is the focus of the present paper. The fragments were collected in various archeological sites in the state of Mato Grosso do Sul at Lalima village. They are thought to be originating from three Indian tribes: the Guarani (66%), the Jacadigo (22%) and the Kadiweu (12%). We use information contained in an archeological researcher's database. It contains qualitative and quantitative observations obtained from the characteristics of the pieces. The researcher's expertise provided precise classification of about 760 pieces. A supervised model of classification was created to infer the Indian technological traditions of the 2300 pieces of fragments collected from the same sites. Bayesian nets were the basis for building the model. Bayesian nets are directed acyclic graphs that properly represent the dependency within a set of random covariates. This kind of network represents the joint probability distribution of these variables and a particular factorization of it. Our approach provides a robust classification: it is based on the probabilities of fragment being originated from each one of the three archeological communities. Also, if the probability of technological tradition indicates "low probabilities" for all three groups, there could be an indication of the presence of an additional community. Comparison with alternative methods to build the networks was also presented.