Abstracts

Asymptotic properties of L_1 estimators in a multi-stage dose-response model: A Monte Carlo study

Carmen D.S. André, *University of São Paulo, Brazil*
Subhash C. Narula, *Virginia Commonwealth University, Richmond, USA*

Abstract: The single-stage and two-stage dose response models are frequently used in practical applications. The maximum likelihood and the least squares principles are often used to estimate the unknown parameters of the model. It has been shown that these methods are sensitive to outliers in the data. The minimum sum of absolute errors MSAE (or L_1) criterion is more resistant to outliers than these popular procedures. However, at present not much is known about the statistical properties of the MSAE estimators of the parameters of the multistage dose-response model. In this paper, our objective is to study asymptotic properties and distribution of the MSAE estimators of the single-stage and two-stage dose-response models by simulation and to find the smallest sample size for which we may use the asymptotic distribution to draw statistical inferences about the parameters. We also give an approximate expression for the variance of these estimators when their asymptotic distribution follows a multinormal distribution.

The L^1-norm and interlaboratory tests

P. L. Davies, *University of Essen, Germany*

Abstract: The form of interlaboratory test we consider is that where each of I laboratories returns exactly one reading for each of J sanples. Such a test may be described by the random effects model

$$X_{ij} = X_i + a_j + \epsilon_{ij}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J.$$

The X_i represent the laboratory effects, the a the sample conaminations and the ϵ_{ij} the measurement errors. The problem is to identify outlying observations and outlying laboratories. As we have only one observation per cell it is commonly believed that it is not possible to detect outliers or, equivalently, non-additivity. As shown in Terbeck and Davies (1996) this is not correct and so called unconditionally identifiable outlier patterns may be found by the L^1- or an appropriate M-functional. The results

of Terbeck and Davies are improved in certain respects and then applied to the random effects model. The method is applied to a real data set considered by Lischer (1993) which is concerned with the determination of lead in sewage sludge.

Multivariate L_1 mean

Yadolah Dodge and Valentin Rousson,
University of Neuchâtel, Switzerland

Abstract: The center of a univariate data set $\{x_1, ..., x_n\}$ can be defined as the point μ that minimizes the norm of the vector of distances $\mathbf{y}' = (|x_1 - \mu|, ..., |x_n - \mu|)$. As the median and the mean are the minimizers of respectively the L_1- and the L_2-norm of $\mathbf{y}$, they are two alternatives to describe the center of a univariate data set. The center $\boldsymbol{\mu}$ of a multivariate data set $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ can also be defined as minimizer of the norm of a vector of distances. In multivariate situations however, there are several kinds of distances. In this paper, we consider the vector of L_1-distances $\mathbf{y}_1' = (||\mathbf{x}_1 - \boldsymbol{\mu}||_1, ..., ||\mathbf{x}_n - \boldsymbol{\mu}||_1)$ and the vector of L_2-distances $\mathbf{y}_2' = (||\mathbf{x}_1 - \boldsymbol{\mu}||_2, ..., ||\mathbf{x}_n - \boldsymbol{\mu}||_2)$. We define the L_1-median and the L_1-mean as the minimizers of respectively the L_1- and the L_2-norm of $\mathbf{y}_1$; and then the L_2-median and the L_2-mean as the minimizers of respectively the L_1- and the L_2-norm of $\mathbf{y}_2$. In doing so, we obtain four alternatives to describe the center of a multivariate data set. While three of them have been already investigated in the statistical literature, the L_1-mean appears to be a new concept. Contrary to the L_1-median, the L_1-mean is proved to be unique in almost all situations. In order to compare these multivariate medians and means, we use the rule of the net advantage coefficient introduced by Stavig and Gibbons (1977). A simulation study shows that the L_1-mean performs well, especially for data sets drawn from bivariate Laplace distribution.

The information for the direction of dependence in L_1 regression

Yadolah Dodge, *University of Neuchâtel, Switzerland*
Joe Whittaker, *Lancaster University, U.K.*

Abstract: An L_1 regression model for a response variable X_2 is to suppose that the conditional distribution of X_2 given X_1 is Laplace, and that the marginal distribution of the explanatory variable X_1 is also Laplace.

We show that there is information to distinguish the direction of dependence X_1 and X_2; or equivalently to distinguish between the models in which X_1 is dependent on X_2, and X_2 is dependent on X_1. This is not true for L_2 regression based on the Normal distribution.

Dimension choice for sliced inverse regression based on ranks

Louis Ferré, *Université Paul Sabatier, Toulouse, France*

Abstract: Sliced Inverse Regression is a method for reducing the dimensionality in multivariate non parametric regression problems. While the selection of the dimensionality has been investigated for the original version, no solution has been proposed for Hsing and Carroll (1992) approach based on order statistics and associated concomitant variables. By using model selection approaches, we propose here two ways for selecting the dimensionality by estimating a loss function: first, a direct estimation is proposed and, then a Jack-Knifed estimate is investigated. Finally, the rank version is compared to classical SIR on a real life data set.

L_1-norm and L_2-norm methodology in cluster analysis

Allan D. Gordon, *University of St Andrews, North Haugh, Scotland*

Abstract: An overview is provided of the use of L_1-norm and L_2-norm methodology in cluster analysis. Topics covered include dissimilarity measures, partitions, fuzzy classifications, hierarchical classifications, and consensus classifications.

Some issues in the applications of conditional quantile functions

Xuming He, *University of Illinois, USA*

Abstract: Conditional quantile functions are useful in a variety of applications. Regression quantiles for linear models have been recently extended to semiparametric and nonparametric models. Further investigations are needed for both the statistical theory and computations. In this paper, I attempt to raise two questions that I believe are important to build a

solid foundation for the applications of quantile regression. They focus on nearly extreme quantiles and the problem of crossing in estimated quantile functions.

The median function on structured metric spaces

F. R. McMorris, *University of Louisville, USA*

Abstract: When (X, d) is a finite metric space and $\pi = (x_1, \ldots, x_k) \in X^k$, a median for π is a element x of X for which $\sum_{i=1}^{k} d(x, x_i)$ is minimum. The function that returns the set of all medians for any tuple π is called the median function on X. A brief survey is given of some of the results concerning the median function, starting with an arbitrary metric space and finishing with the case where X is a set of hypergraphs and d is the metric based on the L_1-norm. A simplistic maximum likelihood interpretation for the median function is also given.

Least absolute value estimation of a linear functional model

Édina Shisue Miazaki, *Statistika Consultoria, Campinas, Brazil*
Gabriela Stangenhaus, *Universidade de Brasília, Brazil*

Abstract: This paper presents two robust L_1 based estimators for the parameters of a simple linear functional relationship (SLFR). The maximum likelihood estimation when the errors follow a double exponencial distribution and the weighted L_1 estimation are solved as non-linear optimization problems. The least median of squares estimates are proposed as starting values and the scale measures of the errors are based on the MAD. Both methods are resistant to outlying observations and the weighted L_1 estimator is resistant to leverage points. Real examples illustrate the methods.

ANOVA - models: A Bayesian analysis

Wolgang Polasek and Shuangzhe Liu, *University of Basel, Switzerland*

Abstract: The 1-way and 2-way ANOVA are formulated as Bayesian linear models with conjugate prior distributions. The classical case is treated

as a special one using matrix generalized (g-) inverses leading to so-called OLS^- and OLS^+ estimates of the rank deficient ANOVA model. The 2-way ANOVA model without interactions can also be estimated in a 2-step procedure.

Robustifying growth curve model estimation

Gabriela Stangenhaus, *Statistika Consultoria, Campinas, Brazil*
Elisete C. Quintaneiro Aubin, *Universidade de São Paulo, Brazil*

Abstract: A robustified version of the parameter matrix estimators in the standard growth curve model obtained via the Potthoff-Roy transformation is presented. The asymptotic distribution of the robust estimators is derived and the estimation of their variance-covariance matrix is discussed.

Fitting L_2 norm classification models to complex data sets

Maurizio Vichi, *University "G.DAnnunzio" di Chieti, Pescara, Italy*

Abstract: In this paper methodologies for fitting classification models (dendrograms and partitions) to two and three-way arrays of dissimilarities minimizing a L_2 norm loss function are examined. A new algorithm for fitting several hierarchical classifications to quite large three-way arrays is also discussed.

Applications of mathematical programming in L_1-estimation of nonlinear models

Jinde Wang, *Nanjing University, China*

Abstract: In this paper we review the results in L_1-estimation of nonlinear models obtained by applying mathematical programming techniques. We describe briefly the ways to find the asymptotic distribution, the approximate representation and to treat dependent random error cases and inequality-constrained cases. With these results one can conclude that mathematical programming is a suitable tool for studying L_1-estimation problems.

Some contributions to M-estimation in regression models

Lincheng Zhao,
University of Science and Technology of China, Hefei, China

Abstract: In this paper, we briefly survey some contributions to asymptotic theory on M-estimation in a linear model and least absolute deviations (LAD) estimation in a censored regression model (known as the Tobit model), as well as on the relevant test criteria in ANOVA in the above models. As a general approach on statistical data analysis, asymptotic theory of M-estimation in regression models has received extensive attention. In recent years, the author and some of his cooperators worked on this field and obtained some new results. In this paper we briefly introduce some of them and the related work in the literature. As a special case, the minimum L_1-norm (ML_1N) estimation, also known as the least absolute deviations (LAD) estimation, plays an important role and is of special interest. Considering this point, we will pay much attention to them as well. Our topics concern the usual linear model and a censored regression model, known as the Tobit model in econometric research.