THE SPEAKERS OF THE WORKSHOP

Invited Speakers

Contributed Speakers

INVITED TALKS

Semiparametric Models in Survival Analysis and Quantile Regression

Probal Chaudhuri[1]

[1]Theoretical Statistics and Mathematics Division, Indian Statistical Institute, Calcutta, India

Abstract

Many of the popular regression models used in survival analysis including Cox's proportional hazard model can be viewed as semiparametric models having some intrinsic monotonicity properties. One is interested in estimating and drawing inference about a finite dimensional Euclidean parameter in that model in the presence of an infinite dimensional nuisance parameter. These survival analysis models are special cases of monotone single index model used in econometrics. The use of average derivative quantile regression techniques for parameter estimation in such models will be discussed. In addition to regression models with univariate response and a single index, we will also discuss possible extensions of the methodology for multivariate response and multiple index models.

Heteroscedastic and autocorrelation consistent estimators of standard errors in robust regression

Christophe Croux[1]

[1]Department of Applied Economics, Katholieke Universiteit Leuven, Leuven, Belgium

Abstract

A regression estimator is said to be robust if it is still reliable in the presence of outliers. On the other hand, its standard error is said to be robust if it is still reliable when the regression errors are autocorrelated and/or heteroscedastic. One speaks about heteroscedastic and autocorrelation consistent (HAC) estimators of standard errors This paper shows how robust standard errors can be computed for several robust estimators of regression. The improvement relative to non-robust standard errors is illustrated by means of large-sample bias calculations, simulations, and a real data example. It turns out that non-robust standard errors of robust estimators may be severely biased. However, if autocorrelation and heteroscedasticity are absent, non-robust standard errors are more efficient than the robust standard errors that we propose. We therefore also present a test of the hypothesis that the robust and non-robust standard errors have the same probability limit.

Shape Constraints and Multiscale Methods for Density Estimation

Lutz Dümbgen[1]

[1]Department of Mathematical Statistics and Actuarial Science, University of Bern, Bern, Switzerland

Abstract

In nonparametric curve estimation, shape constraints such as monotonicity or convexity are known to yield estimators adapting to unknown smoothness properties of the underlying curve. This talk discusses a particular shape constraint in the context of density estimation. We assume that the underlying density is log-concave and show that the resulting nonparametric estimators of the density and distribution function have various interesting properties.

While log-concavity is a reasonable assumption in connection with homogeneous populations, another task in density estimation is inference about modality and local log-concavity or -convexity. We present some multiscale procedures for these purposes yielding confidence statements with guaranteed level for finite samples.

This is joint work with Kaspar Rufibach (Bern) and Guenther Walther (Stanford).

An Affine-Invariant Data Depth Based on Random Hyperellipses

Ryan Elmore[1]

[1]Mathematical Sciences Institute, The Australian National University, Canberra, Australia

Abstract

One of the fundamental concepts in the study of statistical depth is that the depth function should be invariant to the choice of coordinate system. Although this notion of affine-invariance is desirable, most of the current depth functions which satisfy this property are difficult to compute in high dimensions. In this paper, a statistical depth function based on random hyperellipses is proposed which is both affine-invariant and simple to compute in any practical dimension. We will discuss the theoretical properties of the depth measure and outline some of its potential applications. Several examples are presented in order to illustrate these concepts. This is work with Bruce Brown, Tom Hettmansperger, and Fengjuan Xuan.

Semiparametrically Efficient One-Step R-Estimation

Marc Hallin[1], and Davy Paindaveine[1]

[1]Mathematics Department, ISRO, and ECARES, Université Libre de Bruxelles, Bruxelles, Belgium

Abstract

Despite a long history, R-estimation methods, unlike rank tests, never made their way to applications. And, even among the experts of rank-based methods, a pretty widespread opinion is that "ranks are fine for testing but not for estimation".

The reasons for this lack of symmetry between estimation and testing are twofold. Practical reasons first: unlike rank test statistics, R-estimators in general are not given under explicit closed forms, but follow from unpleasant optimization procedures, involving discrete-valued objective functions. More fundamental reasons, too: consistency and asymptotic normality proofs are rather elaborate, and restricted to some traditional cases. And, asymptotic variances of R-estimators typically depend

on the unknown underlying density. Such variances cannot be computed exactly, and cannot be estimated easily. As a result, R-estimators, contrary to rank tests are seldom considered in practice.

Statistical decision theory however suggests that the advantages of rank-based methods depend on the local properties of the model under study, not on the specific inference problem under consideration. From the point of view of the asymptotic theory of statistical experiments, once a type of scores (Wilcoxon, van der Waerden, Laplace, ... ; in general, this choice is associated with some "target density" f) has been chosen, the rank transformation simply consists in mapping the original sequence of statistical experiments $\mathcal{E}^{(n)}$, say, onto another sequence $\mathcal{E}^{(n)}_{f,g}$, where g is the actual unknown density of the observations. The local properties of the resulting model are considered attractive from the point of view of hypothesis testing (distribution-freeness and invariance, local powers). These properties belong to the corresponding local Gaussian shift experiments, hence are fully characterized by $\mathcal{E}^{(n)}_{f,g}$'s information matrices. Typically, the performance of optimal estimators in such models is measured by a covariance matrix which is the inverse of the information matrix characterizing the noncentrality parameter, under local alternatives, of the chi-square distributions of optimal test statistic. If these information matrices are attractive from the point of view of hypothesis testing, *they should be equally attractive from the point of view of point estimation.* Actually, it has been shown by Hallin and Werker (2002) that, under very general assumptions, these matrices coincide with the semiparametrically efficient information matrices.

Now, the practical problems related with the implementation of R-estimation remain. In a sense, they are the same as for the implementation of most M-estimators, including the maximum likelihood ones: as a rule, no explicit form is provided, and the estimator results from the minimization (maximization) of some rank-based objective function. In R-estimation, however, the form of objective functions, which are intrinsically piecewise constant, creates some additional trouble. To the best of our knowledge, the problems resulting from this discrete nature of rank-based objective functions have been solved for location and linear regression models only (the seminal paper in this direction is Jurečková 1971). An unsuccessful attempt has been made in the context of linear (ARMA) time-series models (Allal et al. 2001), who explain why the classical rank-based objective function approach fails in that case.

The usual way to escape numerical optimization, and to provide closed form versions of asymptotically optimal estimators is (in the context of LAN experiments) Le Cam's *one-step* method. Implementation of this method in the present context however runs into the same difficulties as the estimation of asymptotic covariance matrices of R-estimators. We are showing how a very intuitive local maximum likelihood argument allows for a simple and feasible solution.

Applications include classical R-estimation procedures but also less traditional ones, such as those involving serial rank statistics, rank-and-sign statistics, or multivariate, hyperplane-based signed rank statistics.

Bayesian R-Estimates

Tom Hettmansperger[1], and Xiaojiang Zhan[2]

[1]Department of Statistics, The Pennsylvania State University, State College, PA, USA
[2]Merck & Co.

Abstract

When prior information exists, it may be desirable to incorporate it in a data analysis, even when we are using robust rank-based methods. In this talk we discuss the implementation of nonparametric rank-based procedures in a Bayesian context. We summarize the information in a sample of data via the (possibly asymptotic) distribution of some rank-based quantity, and use that distribution as a pseudo-likelihood. Meanwhile, we suppose a prior distribution for the parameter(s) of interest in the model. By Bayes' theorem, we can obtain the complete posterior distribution (or the posterior distribution up to a normalizing constant) of the parameter(s) given the rank-based quantity. Statistical inference then proceeds based on this posterior distribution. The one-sample location model is considered using several rank-based quantities from common scores statistics such as the sign statistic, the Wilcoxon signed rank statistic and the normal scores statistic.

Nonparametric Methods and Extreme Value Analysis

Jürg Hüsler [1]

[1]Department of Mathematical Statistics and Actuarial Science, University of Bern, Bern, Switzerland

Abstract

We discuss some nonparametric ideas in the class of extreme value distributions and extreme value models. The class of extreme value distributions are applied in the analyses of extreme values in finance, ecology and other fields. The class is a three parameter family of asymmetric distributions which have some further interesting features, as e.g. heavy tails and finite endpoints of the support. The aim of this work

is to evaluate whether the nonparametric methods can be advantageous also in this class of distributions.

Asymptotics for Extreme Regression Quantiles

Jana Jurečková [1]

[1]Department of Probability and Statistics, Charles University in Prague, Prague, Czech Republic

Keywords: Regression Quantile, R-estimator, Extreme Regression Quantile.

Abstract

Consider the linear regression model

$$\mathbf{Y} = \beta_0 \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{E} \tag{1}$$

with observations $\mathbf{Y} = (Y_1, \ldots, Y_n)'$, i.i.d. errors $\mathbf{E} = (E_1, \ldots, E_n)'$ with an unknown distribution function F, increasing on the set $\{x : 0 < F(x) < 1\}$, and unknown parameter $\boldsymbol{\beta}^* = (\beta_0, \beta_1, \ldots, \beta_p)'$. The extreme (maximal) regression quantile is defined as a solution of the linear program $\sum_{i=1}^n (b_0 + \mathbf{x}_i'\mathbf{b}) =:$ min under the restrictions $b_0 + \mathbf{x}_i'\mathbf{b} \geq Y_i,\ i = 1, \ldots, n,\ b_0 \in \mathbf{R},\ \mathbf{b} \in \mathbf{R}^p$. Jurečková and Picek (2005) showed that the extreme regression quantile can be equivalently written in a two step version, starting with an R-estimator $\widetilde{\boldsymbol{\beta}}_{nR}$ of the slope parameters, generated by the score function $\varphi(u) = I[u \geq 1 - \frac{1}{n}] - \frac{1}{n},\ 0 \leq u \leq 1$, and then ordering the residuals with respect to $\widetilde{\boldsymbol{\beta}}_{nR}$. Jurečková (2005) showed that, provided the density f of the E_i belongs to the domain of attraction of the Gumbel extreme distribution and $nf(F^{-1}(1 - \frac{1}{n})) \to \infty$ as $n \to \infty$, the slope component $\widetilde{\boldsymbol{\beta}}_{nR}$ of the extreme regression quantile consistently estimates $\boldsymbol{\beta}$ and admits the asymptotic representation

$$\begin{aligned} &nf\left(F^{-1}(1 - \tfrac{1}{n})\right)\left[\widetilde{\boldsymbol{\beta}}_{nR}(1 - \tfrac{1}{n}) - \boldsymbol{\beta}\right] \qquad (2) \\ &= n\Big(\sum_{i=1}^n (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)'\Big)^{-1} \sum_{j=1}^n (\mathbf{x}_{nj} - \bar{\mathbf{x}}_n) \\ &\quad \left[a_n(R_j(\mathbf{0}), 1 - \tfrac{1}{n}) - (1 - \tfrac{1}{n})\right] + o_p(1)\Big[= O_p(1)\Big] \end{aligned}$$

where $R_j(\mathbf{0})$ is the rank of Y_i among $Y_1, \ldots, Y_n$ under $\boldsymbol{\beta} = \mathbf{0}$, $\bar{\mathbf{x}}_n = n^{-1}\sum_{i=1}^n \mathbf{x}_{ni}$ and

$$a_n(j,\alpha) = \begin{cases} 0, & j \leq n\alpha, \\ j - n\alpha, & n\alpha \leq j \leq n\alpha + 1, \\ 1, & n\alpha + 1 \leq j, \quad j = 1, \ldots, n. \end{cases}$$

are Hájek's rank scores. If $\mathbf{x}_{n1}, \ldots, \mathbf{x}_{nn}$ are random, independent of $E_1, \ldots, E_n$, and create a random sample from a p-variate distribution function H with expectation $\mathbf{0}$ and satisfying $\frac{1}{n}\sum_{i=1}^n \mathbf{x}_{ni}\mathbf{x}_{ni}' \xrightarrow{p} \mathbf{Q}$ as $n \to \infty$, with a positively definite matrix $\mathbf{Q}$ of order $p \times p$, then the representation (2) changes to the form

$$nf\left(F^{-1}(1-\tfrac{1}{n})\right)\left[\widetilde{\boldsymbol{\beta}}_{nR}(1-\tfrac{1}{n}) - \boldsymbol{\beta}\right] \tag{3}$$

$$= \mathbf{Q}^{-1}\sum_{j=1}^n \mathbf{x}_{nj}\left[a_n(R_j(\mathbf{0}), 1-\tfrac{1}{n}) - (1-\tfrac{1}{n})\right] + o_p(1)\Big[= O_p(1)\Big].$$

The representations (2) and (3) enable to derive the asymptotic distributions of $\left\{nf\left(F^{-1}(1-\frac{1}{n})\right)\left[\widetilde{\boldsymbol{\beta}}_{nR}(1-\frac{1}{n}) - \boldsymbol{\beta}\right]\right\}_{n=1}^{\infty}$ both for the random and nonrandom $\mathbf{x}_{ni}$.

Acknowledgement

The research was partially supported by the Czech Republic Grant 201/05/2340 and by the Research Project MSM 0021620839.

References

J. Hájek (1965). Extension of the Kolmogorov-Smirnov test to regression alternatives. *Proc. of Bernoulli-Bayes-Laplace Seminar* (L. LeCam, ed.), pp. 45–60. Univ. of California Press.

J. Jurečková (2005). Regression Quantiles and Hájek's Rank Scores. *ICORS'2005* (abstract).

J. Jurečková and J. Picek (2005). Two-step regression quantiles. Submitted.

S. Portnoy and J. Jurečková (1999). On extreme regression quantiles. *Extremes*, 2:3, 227–243.

An Empirical Comparison of Ensemble Methods Based on Classification Trees

Mounir Hamza[1] and Denis Larocque[1]

[1]Department of Quantitative Methods, HEC Montréal, Montréal, Québec, Canada

Keywords: Bagging, Boosting, Arcing, Random forest, Classification tree, CART, Noise, Linear combination of variables, Splitting rule, Gini, Entropy, Twoing.

Abstract

In this paper, we perform an empirical comparison of the classification error of several ensemble methods based on classification trees. This comparison is performed by using fourteen data sets that are publicly available and that were used in Lim, Loh and Shih (Machine Learning 40, 203-228, 2000). The methods considered are a single tree, Bagging, Boosting (Arcing) and random forests. They are compared from different perspectives. More precisely, we look at the effects of noise and of allowing linear combinations in the construction of the trees, the differences between some splitting criteria and, specifically for random forests, the effect of the number of variables from which to choose the best split at each given node. Moreover, we compare our results with those obtained in Lim et al. (2000). In this study, the best overall results are obtained with random forests. In particular, random forests are the most robust against noise. The effect of allowing linear combinations and the differences between splitting criteria are small on average, but can be substantial for some data sets.

Mining Massive Text Data and Developing Robust Statistical Tracking

Regina Liu[1]

[1]Department of Statistics, Rutgers University, Piscataway, NJ, USA

Abstract

We present a systematic data mining procedure for exploring large free-style text datasets to discover useful features and develop tracking statistics, often referred to as performance measures or risk indicators. The procedure includes text classification,

construction of tracking statistics, inference under error measurements. An aviation safety report repository from the FAA is used to illustrate applications of our research to aviation risk management and general decision-support systems. Some specific text analysis methodologies and tracking statistics are discussed. A robust approach for incorporating misclassified data or error measurements into the inference for tracking statistics is proposed and evaluated.

Although most illustrations here are drawn from aviation safety data, the proposed data mining procedure with its robust inference framework applies to many other domains, including, for example, mining free-style medical reports for tracking possible disease outbreaks.

This is joint work with Daniel Jeske, Department of Statistics, UC Riverside.

Rank-Based Analyses of Multivariate Linear Models with Applications to Profile Analysis

Joseph McKean[1], John Kloke[1], and Majda Salaman[1]

[1]Department of Mathematics and Statistics, Western Michigan University, Kalamazoo, MI, USA

Abstract

In this talk, we present several new rank-based procedures for the analysis of multivariate linear models. One procedure is an affine equivariant estimate for the regression coefficient matrix of the multivariate linear model. These estimates are based on a transformation and retransformation technique that uses Tyler's (1987) M-estimator of scatter. The proposed estimates are obtained by retransforming the componentwise rank-based estimate due to Davis and McKean (1993) and a componentwise generalized rank estimate. This procedure is for the general linear multivariate model. For repeated measure type responses, we discuss a rank-based GEE procedure and a rank-based procedure which utilizes Arnold's (1981) initial transformation of the responses. Asymptotic theory is presented for all the procedures. We then compare the methods over a simulation study on profile models.

The Place of Depth in Nonparametric Statistics

Ivan Mizera[1]

[1]Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada

Abstract

Although halfspace depth was originally proposed in the context of sign test for bivariate data, and also as a possible multivariate ranking device, the subsequent, and relatively recent development of its generalizations to various data-analytic situations somehow did not emphasize the connection to classical methods. We would like to offer some views on the possible use of halfspace depth in statistical analyses performed in a nonparametric way and point out some specific virtues and problems of this type of approach. A special focus will be on data coming from controlled experiments that are typically analyzed via linear model techniques.

Rank-Based Inference on the Shape of Elliptical Distributions

Marc Hallin[1], and Davy Paindaveine[1]

[1]Mathematics Department, ISRO, and ECARES, Université Libre de Bruxelles, Bruxelles, Belgium

Abstract

We propose (i) a class of rank-based procedures for testing that the shape matrix of an elliptical distribution is equal to some fixed value (this problem includes the problem of testing for sphericity as a particular case), as well as (ii) a class of R-estimators for the shape parameter. The proposed tests/estimators are invariant/equivariant under translations, monotone radial transformations, rotations, and reflections with respect to the estimated center of symmetry. For adequately chosen scores, they are locally asymptotically optimal (in the Le Cam sense) at given densities. The multivariate ranks used throughout are those of the distancesin the metric associated with the null value of the shape matrix (for testing problems) or with a preliminary estimate of the shape parameter (for the estimation problem)between the observations and the estimated center of the distribution. Asymptotic relative efficiencies with respect to the standard Gaussian procedures (i.e., pseudo-Gaussian LRT and MLE) are derived, and are shown to be uniformly high for specific choices of the score functions. The proposed tests are valid without any moment assumption. As for the proposed R-estimators, they are defined as iterative M-estimators. Unlike those obtained via the methods described in Marc Hallins talk, these do not require the difficult estimation of a cross-information coefficient. Nevertheless, they are root-n consistent only under a (very) mild condition on this unknown cross-information coefficient. We also compute their influence functions and show that, similarly as for univariate R-estimation for location, a broad range of robustness behaviors can be obtained by considering various types of score functions.

On Zeros in the Sign and Signed Rank Test

Ronald H. Randles[1]

[1]Department of Statistics, University of Florida, Gainesville, FL, USA

Abstract

The traditional method of deleting zero observations in the sign and signed rank tests are in many applications, the right answer to the wrong question. In these settings the zeros should play a role that favors the null hypothesis. This talk will emphasize ways to use the zeros in a conservative manner, but one that produces good power for the appropriate question. Particular emphasis is placed on two-tailed test settings, because a method of obtaining an appropriate p-value in these cases is less obvious.

The Spatial Multivariate Quantile Function: Strengths, Weaknesses, Competitors

Robert Serfling[1]

[1]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, TX, USA

Abstract

In recent years a number of quite different multivariate depth and quantile functions have been formulated and investigated. A timely question is whether one of these can or should be adopted as the best choice for practical use. This talk will focus on the spatial multivariate quantile function (Choudhuri, Dudley) as a benchmark. For this case, we first will review basic features and characterizations and associated depth, centered rank, and outlyingness functions. We then will treat productive results now available for this quantile function: the influence function, masking and swamping breakdown points of associated outlier identification procedures, cluster analysis procedures, multivariate sign test procedures, and a Bahadur-Kiefer (B-K) representation. We also will introduce an extension to a spatial U-quantile function, along with an extended B-K theorem. For the empirical spatial U-quantile function, the B-K representation provides a useful U-statistic approximation. In terms of spatial U-quantiles, interesting new multivariate nonparametric estimators and test statistics can be formulated, for example generalized multivariate signed-rank tests, an extension to multiple regression of Theil's nonparametric simple linear regression slope estimator, and a new matrix-valued dispersion measure whose sample analogue estimator has breakdown point 0.293 independently of the dimension of the data. We will also discuss equivariance limitations of the spatial quantile function and explore whether these can be overcome by a suitable modification. Finally, we will examine major competing depth and quantile approaches comparatively against the strengths and weaknesses of the spatial quantile function.

Bandwidth Selection for Kernel Density Estimates Based on Data Sharpening

Simon Sheather[1]

[1]Department of Statistics, Texas A&M University, College Station, TX, USA

Abstract

A new general method for reducing bias in density estimation has been proposed by Hall and Minnotte (2002, JRSSB). The method is known as data-sharpening since it involves moving the data away from regions where they were sparse towards regions where the density is higher. Once the data have been "sharpened" they are used in a kernel estimator to produce a less biased estimator. In this talk, we shall consider the problem of choosing the bandwidth for sharpened density estimates.

A Nonparametric Multivariate Multisample Test

Shoja'eddin Chenouri[1], and Christopher G. Small[2]

[1]School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada

[2]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

Abstract

In this talk, we shall consider a family of nonparametric multivariate multisample tests based on depth rankings. These tests are of Kruskal-Wallis type in the sense that the samples are variously ordered. However, unlike the Kruskal-Wallis test, these tests are based upon a center-outward ranking using a statistical depth function such as the halfspace depth or the Mahalanobis depth, etc. Unlike the univariate case, multivariate data sets can be ordered using many different depth-based orderings. The types of tests we propose are adapted to the depth function that is most appropriate for the application. Under the null hypothesis that all samples come from the same distribution, we show that the test statistic asymptotically has a chi-square distribution. In addition, for small sample sizes, the test statistic is exactly distribution-free. Some comparisons of power are made with the Hotelling T^2, and the test of Choi and Marden (1997). Our test is particularly recommended when the data are of unknown distribution type where there is some evidence that the density contours are not elliptical. However, when the data are normally distributed, we often obtain relative power over 95%.

Invariant coordinate selection (ICS): A nonparametric view of independent components analysis (ICA)

David E. Tyler[1]

[1]Department of Statistics, Rutgers University, Piscataway, NJ, USA

Abstract

An obvious method for generating measures of location for p-dimensional distributions is to simply apply univariate measures of location to each of the coordinates, e.g. the coordinatewise median. A drawback to this approach is that the resulting measure of multivariate location is not affine equivariate. If one could select the coordinates in an invariant manner, however, i.e. select p data dependent linear combinations of the variables which are invariant under nonsingular transformations of the variables, then applying coordinatewise measure of univariate location to the transformed variables and then back-transforming gives an affine equivariant measure of multivariate location. Affine covariant measures for the scatter matrix can also be generated using coordinatewise measures of scale.

To be more specific, let $Y = \{y_1, \dots, y_n\}$ be a p-dimensional data set. Suppose we are able to define a nonsingular matrix $A(Y)$ such that the transformed p-dimensional data set $Z = A(Y)Y$ is invariant under nonsingular transformations of Y, i.e. $A(Y)Y = A(BY)BY$ for any nonsingular matrix B. If we then apply univariate measures of location and scale to each of the components of Z producing $\mu(Z) \in \Re^p$ and $\sigma(Z) \in \Re^p$ respectively, then affine equivariant measures of multivariate location and scatter can be defined by

$$\mu(Y) = A(Y)^{-1}\mu(Z) \quad \text{and} \quad \Sigma(Y) = A(Y)^{-1}D(\sigma^2(Z))(A(Y)')^{-1},$$

where $D(\cdot)$ is a diagonal matrix whose diagonal elements are given by its vector argument.

One method for generating such an invariant transformation is as follows. First compute two different affine covariate estimates of scatter for Y, say V_o and V_1, and then define $A(Y) = (a_1, \dots, a_p)$ such that

$$V_o a_j = \gamma_j V_1 a_j \quad \text{for} \quad j = 1, \dots, p \quad \text{or equivalently,} \quad V_o A(Y) = V_1 A(Y)\Delta,$$

where $\Delta = D(\gamma_1, \dots, \gamma_p)$. That is, $A(Y)$ are the principal component vectors of V_o, relative to the Mahalonobis inner product defined via V_1. The transformed variates $Z = A(Y)Y$ can be viewed as *affine invariant principal components.* In a personal communication, Hannu Oja has noted that under certain conditions, the matrix $A(Y)^{-1}$ also represents a solution to the independent component analysis problem.

Beside using the transformed variate Z to generate measures of multivariate location and scatter, these transformed variates can also be used to generate multivariate generalizations of univariate concepts, e.g. affine equivariate quantiles. They can

also be used for generating affine invariant nonparametric tests, e.g. an affine invariant sign tests which is asymptotically nonparametric over the class of all symmetric multivariate distributions and not just elliptically symmetric distributions.

Finally, we note that one can produce affine invariant diagnostic plots by plotting the components of Z or by making pairwise scatter plots of the components of Z. We give several examples which illustrates the utility of the proposed methods.

This talk is based on joint work with Oja Hannu of the University of Jyväskylä and Lutz Dümbgen of the University of Bern.

Multi-Dimensional Trimming Based on Projection Depth

Yijun Zuo[1]

[1]Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA

Abstract

With a very natural order principle, trimming in one dimension is straightforward. Univariate trimmed means are among most popular estimators of location parameters. They can be very robust against outliers and heavy tailed distributions while enjoying a very high efficiency at a variety of distributions. Multi-dimensional data often contain outliers and are "heavy tailed". Extending trimming idea to the multi-dimensional setting is quite desirable. The task, however, becomes non-trivial. In this talk, multi-dimensional trimming based on data depth is discussed. It is found that multi-dimensional depth trimmed means can possess very desirable properties such as high efficiency as well as high robustness. Consequently they can serve very well as multi-dimensional location estimators. Trimmed means based on different notions of data depth are also compared based on their performance. Inference procedures based on the depth trimmed means are discussed.

CONTRIBUTED TALKS

Robust Correlation Applied to Locating Landmarks

Jan Kalina[1]

[1]University of Duisburg–Essen, Department of Mathematics, Essen, Germany

Keywords: Computational aspects of robust methods, Human faces.

Abstract

In our work with images of human faces (joint work with P.L. Davies), similarity between two images must be measured in a robust way. Some of suitable correlation measures are based on robust regression (least trimmed squares or least weighted squares), other examples include directly the maximal weighted correlation coefficient over all permutations of the weights.

Such methods are computationally intensive and can be only approximated. We generalize the algorithm of Kalina (2003) to approximate the minimum of the weighted loss function in both regression and correlation context.

In an example we find a better approximation to the least trimmed squares estimator than software packages R and S-Plus. Then we use the methods to automatically search for the vertical axis of symmetry in human faces or to locate the eyes using templates.

References

J. Kalina (2003). Autocorrelated disturbances of robust regression. In Fournier B. et al., eds., *Proceedings EYSM 2003*, pp. 65-72, Ovronnaz, Switzerland.

Graphical Comparison of Multivariate Nonparametric Location Tests for Restricted Alternatives

Michael Vock[1]

[1]Department of Mathematical Statistics and Actuarial Science, University of Bern, Bern, Switzerland

Abstract

There have been several proposals of nonparametric tests for restricted (or "one-sided") multivariate location alternatives. The selection of a suitable test for a specific problem is an open question. We discuss the most common types of hypotheses and present a graphical means of assessing the adequacy of a test for the different types of hypotheses. This leads to a classification of the test procedures. In contrast to a graphical representation using rejection regions (which is frequently used in the parametric context), our approach is suitable for the comparison of tests based on entirely different statistics.