

Symposium Program

Friday, January 22

Room 349, Reitz Union

- 7:45 - 8:20 Breakfast
-

- 8:20 - 10:00 **Session I**

- 8:20 - 9:00

Bootstrap Testing for Ordered Multivariate Means

William Schucany

Southern Methodist University

Many practical problems involve detecting trends. By considering differences in adjacent levels, an increasing sequence of three may be transformed to an equivalent problem of detecting the positive quadrant. The likelihood-ratio test (LRT) of the null hypothesis that a multivariate mean equals zero versus the positive-orthant alternative is reexamined. Perlman (1969) derived the exact null distributions under normality for general cone alternatives. However, because these distributions depend on the unknown covariance matrix, the usable critical points have only been bounds. For important cases the resulting one-sided tests are biased. The disappointing performances of these approximate LRT have been the subject of several critical articles over the years.

The bootstrap can rescue the LRT by estimating the approximate critical point. Monte Carlo comparisons confirm its superiority to Hotelling's T^2 , a "half-space" alternative investigated by Tang (1994), and closely related simple test due to Follmann (1996). Wang and McDermott

(1998) show that their conditional LRT is unbiased and uniformly more powerful than these but they must evaluate integrals numerically. The proposed bootstrap test performs well even if the distribution is not multivariate normal. In addition, it is easy to extend the methodology to general cone-shaped critical regions.

- 9:10 - 9:50

Robust Estimation via the Biased-Bootstrap

Brett Presnell

University of Florida

We introduce a new nonparametric approach to distance-based robust estimation, in which parameters are estimated by evaluating the corresponding functional on a weighted version of the empirical distribution. We call this biased-bootstrap "trimming", although no observations are actually deleted by the procedure. Our approach has two variants, corresponding to two different methods of choosing the weights.

In fixed-dispersion biased-bootstrap trimming, weights are chosen to minimize the distance between the weighted and ordinary empirical distributions, while fixing a measure of the dispersion of the weighted distribution, such as its (generalized) variance. We measure distance by Read and Cressie's power-divergence, though other distances can be used. In practice this approach requires an external robust estimate of scale, but yields a redescending estimator that is competitive with the Huber estimator in location problems.

A dual to the fixed-dispersion approach is fixed-distance trimming. Here a measure of dispersion is minimized while fixing the distance between the weighted and ordinary empirical distributions. Importantly, an appropriate distance can be determined via breakdown considerations, and without reference to the data at hand or its dimension. In the case of location estimation, this yields a robust estimator that requires no external estimate of scale. The method is thus especially useful in multivariate problems where robust estimates are not readily available. In fact, if the measure of dispersion used is the generalized variance,

then the fixed-distance approach yields robust and affine equivariant estimates of both location and scale.

- 10:00 - 10:20 Break
-

- 10:20 - 12:00 **Session II**

- 10:20 - 11:00

(Almost) Nonparametric Inference for Mixture Models

Thomas Hettmansperger
Pennsylvania State University

In this talk we consider ways to estimate the mixing proportions in a finite mixture distribution or estimate the number of components of the mixture distribution without making parametric assumptions about the component distributions. We require a vector of observations on each subject. This vector is mapped into a vector of zeros and ones and summed. The resulting distribution of sums can be modelled as a mixture of binomials. We then work with the binomial mixture. Efficiency and robustness of this method are compared to the strategy of assuming multivariate normal mixtures when, typically, the true underlying mixture distribution is different. It is shown that in many cases the approach based on simple binomial mixtures is superior.

- 11:10 - 11:50

Nonparametric Empirical Bayes Estimation in Semiparametric Linear Models

Pranab K. Sen

University of North Carolina, Chapel Hill

The empirical Bayes methodology, in a parametric setup, incorporates a conjugate prior in the estimation of the Bayes (shrinkage) factor that characterizes the estimator. The situation becomes more complex in semiparametric and (even more in) nonparametric models. Using the Hajek convolution theorem and first-order asymptotic representations for semiparametric and nonparametric estimators it is shown that the Zellner (Gaussian) g-prior on the regression parameters can be readily adopted to formulate suitable empirical Bayes estimators that are essentially related to the robust Stein-rule estimators. Empirical Bayes statistical functionals and adaptive versions are also considered in the same vein.

-
- 12:00 - 1:20 Lunch Break - Arredondo Room, 4th floor of the Reitz Union

-
- 1:20 - 3:00 **Session III**

- 1:20 - 2:00

Estimation with Recurrent Data

Myles Hollander

Florida State University

We consider the problem of nonparametrically estimating the distribution function governing the time to occurrence of a recurrent event based on data accruing from an informative sum-quota stopping rule. Finite-sample and asymptotic properties of the estimators are presented. Furthermore, loss in efficiency is studied for the cases (1) when the right-censored last observation for each unit is ignored and (2) when only the first observation for each unit is used. The procedures are

illustrated using the gastroenterology data set in Aalen and Husebye (1991).

- 2:10 - 2:50

Permutation Tests for Comparing Survival Functions Based on Paired Censored Data with Missing Observations

P. V. Rao

University of Florida

Permutation tests are proposed for comparing marginal survival functions based on three independent randomly right censored samples - a sample of size n from a bivariate lifetime distribution and samples of sizes s and k from the two marginals. Exact tests for testing equality of survival functions based on two independent samples and paired samples can be obtained by setting $n = 0, s, k > 0$ and $n > 0, s = k = 0$, respectively. Since there exist several tests for these two cases, the proposed tests provide new alternatives. However, there is no satisfactory method for comparing survival functions when the data consist of n complete pairs and $s + k$ singletons ($n > 0$ and $s + k > 0$). The proposed tests fill this gap by providing a method for handling this case. The null distribution of the test statistic is asymptotically normal and simulation results show that the tests have good power for detecting scale and location shifts in exponential and log-logistic distributions.

- 3:00 - 4:00 Picture/Break
-

- 4:00 - 5:00 **Session IV - Poster Session**
-

- Dinner - Keene Faculty Center, Dauer Hall

- Featured Speaker - **Robert Hogg**

Saturday, January 23

Room 100, Griffin-Floyd Hall

- 7:45 - 8:20 Breakfast

-
- 8:20 - 10:00 **Session V**

- 8:20 - 9:00

Monte Carlo Evaluation of Resampling-Based Hypothesis Tests

Dennis Boos

North Carolina State University

New resampling-based tests should be evaluated by Monte Carlo simulation, but the resulting nested computations can be overwhelming. A new generalized jackknife procedure is proposed to reduce the size of the inner resampling loop to be much smaller than would be used in the analysis of real data. A version of the method that uses simple linear extrapolation is shown to perform well in correcting for bias and thus reduces overall computation time in Monte Carlo power studies.

-
- 9:10 - 9:50

Exact Bootstrap Moments of an L -estimator

Michael Ernst
University of Florida

Because many bootstrap problems are analytically intractable, the bootstrap is commonly viewed solely as a resampling technique. We show that for the broad class of statistics that are linear combinations of order statistics (L -estimators) exact analytic expressions for the bootstrap mean and variance can be obtained, eliminating the error due to bootstrap resampling. The expressions follow from direct calculation of the bootstrap mean vector and covariance matrix of the whole set of order statistics. We examine the non-negligible error of the resampling approach for estimating the bootstrap variance using some classical L -estimators such as the trimmed mean and the median on some real data and consider an application of these exact estimates in linear regression. We also consider exact percentiles and moments of more general functions of order statistics.

-
- 10:00 - 10:20 Break

-
- 10:20 - 12:00 **Session VI**

- 10:20 - 11:00

Tail Behavior of Extreme Regression Quantiles

Jana Jurecková
Charles University, Prague, Czech Republic

Jureckova (1981) introduced a tail behavior measure of performance of equivariant estimators of location. He, *et al.* (1990) showed that, for a broad class of estimators, this measure is equivalent to the finite sample breakdown point of Huber and Donoho and extended this measure to the linear regression model.

Koenker and Bassett (1978) introduced regression quantiles to generalize the notion of order statistic from the case of a single sample to the linear regression setting. Since then, most of the one-sample properties of quantiles have found natural analogues in the regression quantile situation. The extreme regression quantiles, analogous to the smallest and largest order statistics, were only recently considered in depth by Smith (1994), who developed some asymptotic results for algebraically tailed error distributions.

We shall consider the tail behavior of a linear form of extreme regression quantiles in the linear regression model with regression matrix of order $n \times p$. In the location submodel, this form coincides with the pertaining extreme order statistic. The tail behavior differs depending on whether the basic distribution of errors has heavy (algebraic) or exponential tails and is generally different from that in the location model. For the exponentially tailed distribution, it involves the dimension p , while in the case of heavy tailed distribution it is more similar to the location situation.

-
- 11:10 - 11:50

Efficiency Robust Tests for Survival or Ordered Categorical Data

Joseph Gastwirth

George Washington University

The selection of a single method of analysis is problematic when the data could have been generated by one of several possible models. We show that two efficiency robust approaches both depend on the correlation matrix of the standardized optimum test statistics for each of the models. The first procedure, the MERT, uses the linear combination of these statistics that maximizes the minimum efficiency across all models in the family. The second procedure, the MX, uses the maximum of the optimal statistics. We show that both the MERT and MX approaches can be used to obtain efficiency robust procedures for a wide variety of survival analyses and dose response data, including the combination of ordered tables. We illustrate these with several biomedical data sets. Because the distribution of the maximum of several statistics is more complicated than the corresponding MERT, we provide guidelines

derived from the correlation matrix of the optimum tests to assist the user in deciding when the MX statistic is preferable to the MERT.

- 12:00 - 1:20 Lunch Break - Keene Faculty Center, Dauer Hall
-

- 1:20 - 3:00 **Session VII**

- 1:20 - 2:00

Affine Equivariant Sign and Rank Covariance Matrices

Hannu Oja

University of Jyväskylä, Finland

We consider affine equivariant empirical multivariate sign and rank processes based on the Oja (1983) median. It is shown that the corresponding multivariate sign and rank covariance matrices are then also affine equivariant and carry information about the shape and geometry of the underlying distribution. In the elliptic case, the eigenvalues and eigenvectors of the usual covariance matrix, i.e., the correlation structure, as well as (multivariate multiple) regression coefficients can be estimated using these matrices. Robustness and efficiency properties of the resulting estimates are illustrated with some examples, empirical sensitivity (influence) plots and simulation studies. Finally, tests for multinormality (which compare the sizes, i.e., determinants, of the covariance matrices) are proposed.

- 2:10 - 2:50

Theory and Applications of Data Depth

Regina Liu
Rutgers University

A data depth is a measure of how deep or central a given point is with respect to a multivariate distribution. It gives rise to a new set of parameters for the distribution, which can easily quantify its many complex multivariate features. These parameters can also be visualized by simple graphs. Furthermore, the center-outward ranking of the sample points provided by a data depth leads to a systematic nonparametric inference scheme. We discuss examples in rank tests, multivariate process control charts, constructions of confidence regions, determinations of P-values for a vast range of tests, and regression analysis.

-
- 3:00 - 3:20 Break

-
- 3:20 - 5:00 **Session VIII**

- 3:20 - 4:00

Iterated Multivariate Ranks -- Truly Nonparametric Procedures?

John Marden
University of Illinois

In multiple dimensions, the distribution of the multivariate ranks defined using the multivariate signs depends on distribution of the underlying data, so that procedures based on these ranks are not distribution-free. (They are distribution-free in one dimension.) Several authors have developed multivariate rank-based procedures that are affine invariant and distribution-free under elliptically symmetric distributions.

This talk explores an approach that is conjectured to give distribution-free multivariate ranks under very general conditions. The approach is to take the multivariate ranks, then the multivariate ranks of the ranks, then the multivariate ranks of those, etc., until these ranks converge. The conjecture is that the limiting distribution of these "iterated" ranks is independent of the underlying distribution of the original variables, at least if the underlying variables have a continuous distribution not concentrated on a lower-dimensional plane.

Evidence for the truth of the conjecture arises from taking various data sets and finding the multivariate iterated ranks of the observations. In all examples (as long as the points are not exactly collinear), the iterated ranks appear to converge to a spherically symmetric distribution. Furthermore, when the sample size is not too small, this distribution appears to be approximately the same spherically symmetric one no matter what the initial distribution. For two dimensions, this conjectured limiting distribution has a density over the unit disk that is not uniform but more concentrated at radii around 0.70. In higher dimensions, it appears that the distribution is uniform on the sphere.

If this approach is valid, one can easily perform truly nonparametric analysis in one-way multivariate analysis of variance and regression. Tests of goodness of fit of data to a particular distribution F can be constructed by looking at the iterated ranks relative to the target F . Inverting the iterated rank function would provide a method for simulating observations from F , analogous to the inverse probability transform for univariate distributions.

-
- 4:10 - 4:50

A Simpler, Affine-Invariant Multivariate Distribution-Free Sign Test

Ronald Randles

University of Florida

The assumptions which underly multivariate sign tests are compared in order to concentrate on weak assumptions that still yield a permutation principle for affine-invariant methods. A transformation-retransformation type of multivariate sign test is proposed, which has a distribution-free property similar to that of the interdirection sign test

(Randles(1989)). It is simpler in nature and easier to compute. It can readily be applied to data of any practical dimension. Its performance characteristics are demonstrated.

- Dinner - Mr. Han's Chinese Restaurant