Sparse principal component analysis via regularized low rank matrix approximation

Shen, H. | Huang, J.Z.

Principal component analysis (PCA) is a widely used tool for data analysis and dimension reduction in applications throughout science and engineering. However, the principal components (PCs) can sometimes be difficult to interpret, because they are linear combinations of all the original variables. To facilitate interpretation, sparse PCA produces modified PCs with sparse loadings, i.e. loadings with very few non-zero elements. In this paper, we propose a new sparse PCA method, namely sparse PCA via regularized SVD (sPCA-rSVD). We use the connection of PCA with singular value decomposition (SVD) of the data matrix and extract the PCs through solving a low rank matrix approximation problem. Regularization penalties are introduced to the corresponding minimization problem to promote sparsity in PC loadings. An efficient iterative algorithm is proposed for computation. Two tuning parameter selection methods are discussed. Some theoretical results are established to justify the use of sPCA-rSVD when only the data covariance matrix is available. In addition, we give a modified definition of variance explained by the sparse PCs. The sPCA-rSVD provides a uniform treatment of both classical multivariate data and high-dimension-low-sample-size (HDLSS) data. Further understanding of sPCA-rSVD and some existing alternatives is gained through simulation studies and real data examples, which suggests that sPCA-rSVD provides competitive results. © 2007 Elsevier Inc. All rights reserved.

Testing for equality between two copulas

Volume 100, Issue 3, March 2009, Pages 377-386
Rémillard, B. | Scaillet, O.

We develop a test of equality between two dependence structures estimated through empirical copulas. We provide inference for independent or paired samples. The multiplier central limit theorem is used for calculating p-values of the Cramér-von Mises test statistic. Finite sample properties are assessed with Monte Carlo experiments. We apply the testing procedure on empirical examples in finance, psychology, insurance and medicine. © 2008 Elsevier Inc. All rights reserved.

Construction of asymmetric multivariate copulas

Volume 99, Issue 10, November 2008, Pages 2234-2250
Liebscher, E.

In this paper we introduce two methods for the construction of asymmetric multivariate copulas. The first is connected with products of copulas. The second approach generalises the Archimedean copulas. The resulting copulas are asymmetric and may have more than two parameters in contrast to most of the parametric families of copulas described in the literature. We study the properties of the proposed families of copulas such as the dependence of two components (Kendall's tau, tail dependence), marginal distributions and the generation of random variates. © 2008 Elsevier Inc. All rights reserved.

The multivariate least-trimmed squares estimator

Agulló, J. | Croux, C. | Van Aelst, S.

In this paper we introduce the least-trimmed squares estimator for multivariate regression. We give three equivalent formulations of the estimator and obtain its breakdown point. A fast algorithm for its computation is proposed. We prove Fisher-consistency at the multivariate regression model with elliptically symmetric error distribution and derive the influence function. Simulations investigate the finite-sample efficiency and robustness of the estimator. To increase the efficiency of the estimator, we also consider a one-step reweighted estimator. © 2006 Elsevier Inc. All rights reserved.

Successive direction extraction for estimating the central subspace in a multiple-index regression

Volume 99, Issue 8, September 2008, Pages 1733-1757
Yin, X. | Liu, B. | Cook, R.D.
Successive direction extraction for estimating the central subspace in a multiple-index regression

Yin, X. | Li, B. | Cook, R.D.

In this paper we propose a dimension reduction method for estimating the directions in a multiple-index regression based on information extraction. This extends the recent work of Yin and Cook [X. Yin, R.D. Cook, Direction estimation in single-index regression, Biometrika 92 (2005) 371-384] who introduced the method and used it to estimate the direction in a single-index regression. While a formal extension seems conceptually straightforward, there is a fundamentally new aspect of our extension: We are able to show that, under the assumption of elliptical predictors, the estimation of multiple-index regressions can be decomposed into successive single-index estimation problems. This significantly reduces the computational complexity, because the nonparametric procedure involves only a one-dimensional search at each stage. In addition, we developed a permutation test to assist in estimating the dimension of a multiple-index regression. © 2008 Elsevier Inc. All rights reserved.

Parametric estimation and tests through divergences and the duality technique

Broniatowski, M. | Keziou, A.

We introduce estimation and test procedures through divergence optimization for discrete or continuous parametric models. This approach is based on a new dual representation for divergences. We treat point estimation and tests for simple and composite hypotheses, extending the maximum likelihood technique. Another view of the maximum likelihood approach, for estimation and tests, is given. We prove existence and consistency of the proposed estimates. The limit laws of the estimates and test statistics (including the generalized likelihood ratio one) are given under both the null and the alternative hypotheses, and approximations of the power functions are deduced. A new procedure of construction of confidence regions, when the parameter may be a boundary value of the parameter space, is proposed. Also, a solution to the irregularity problem of the generalized likelihood ratio test pertaining to the number of components in a mixture is given, and a new test is proposed, based on χ^2-divergence on signed finite measures and the duality technique. © 2008 Elsevier Inc. All rights reserved.

Wiener processes with random effects for degradation data

Wang, X.

This article studies the maximum likelihood inference on a class of Wiener processes with random effects for degradation data. Degradation data are special case of functional data with monotone trend. The setting for degradation data is one on which n independent subjects, each with a Wiener process with random drift and diffusion parameters, are observed at possible different times. Unit-to-unit variability is incorporated into the model by these random effects. EM algorithm is used to obtain the maximum likelihood estimators of the unknown parameters. Asymptotic properties such as consistency and convergence rate are established. Bootstrap method is used for assessing the uncertainties of the estimators. Simulations are used to validate the method. The model is fitted to bridge beam data and corresponding goodness-of-fit tests are carried out. Failure time distributions in terms of degradation level passages are calculated and illustrated. © 2008 Elsevier Inc. All rights reserved.

Nonparametric time series prediction: A semi-functional partial linear modeling

Aneiros-Pérez, G. | Vieu, P.

There is a recent interest in developing new statistical methods to predict time series by taking into account a continuous set of past values as predictors. In this functional time series prediction approach, we propose a functional version of the partial linear model that allows both to consider additional covariates and to use a continuous path in the past to predict future values of the process. The aim of this paper is to present this model, to construct some estimates and to look at their properties both from a theoretical point of view by means of asymptotic results and from a practical perspective by treating some real data sets. Although the literature on the use of parametric or nonparametric functional modeling is growing, as far as we know, this is the first paper on semiparametric functional modeling for the prediction of time series. © 2007 Elsevier Inc. All rights reserved.

Maximum likelihood estimation for multivariate skew normal mixture models

http://www.journals.elsevier.com/journal-of-multivariate-analysis/most-cited/
Maximum likelihood estimation for multivariate skew normal mixture models

Volume 100, Issue 2, February 2009, Pages 257-265
Lin, T.I.

This paper provides a flexible mixture modeling framework using the multivariate skew normal distribution. A feasible EM algorithm is developed for finding the maximum likelihood estimates of parameters in this context. A general information-based method for obtaining the asymptotic covariance matrix of the maximum likelihood estimators is also presented. The proposed methodology is illustrated with a real example and results are also compared with those obtained from fitting normal mixtures. © 2008 Elsevier Inc. All rights reserved.

Tails of multivariate Archimedean copulas

Volume 100, Issue 7, August 2009, Pages 1521-1537
Charpentier, A. | Segers, J.

A complete and user-friendly directory of tails of Archimedean copulas is presented which can be used in the selection and construction of appropriate models with desired properties. The results are synthesized in the form of a decision tree: Given the values of some readily computable characteristics of the Archimedean generator, the upper and lower tails of the copula are classified into one of three classes each, one corresponding to asymptotic dependence and the other two to asymptotic independence. For a long list of single-parameter families, the relevant tail quantities are computed so that the corresponding classes in the decision tree can easily be determined. In addition, new models with tailor-made upper and lower tails can be constructed via a number of transformation methods. The frequently occurring category of asymptotic independence turns out to conceal a surprisingly rich variety of tail dependence structures. © 2009 Elsevier Inc. All rights reserved.

The centred parametrization for the multivariate skew-normal distribution

Volume 99, Issue 7, August 2008, Pages 1362-1382
Arellano-Valle, R.B. | Azzalini, A.

For statistical inference connected to the scalar skew-normal distribution, it is known that the so-called centred parametrization provides a more convenient parametrization than the one commonly employed for writing the density function. We extend the definition of the centred parametrization to the multivariate case, and study the corresponding information matrix. © 2008 Elsevier Inc. All rights reserved.

Tail dependence functions and vine copulas

Volume 101, Issue 1, January 2010, Pages 252-270
Joe, H. | Li, H. | Nikoloulopoulos, A.K.

Tail dependence and conditional tail dependence functions describe, respectively, the tail probabilities and conditional tail probabilities of a copula at various relative scales. The properties as well as the interplay of these two functions are established based upon their homogeneous structures. The extremal dependence of a copula, as described by its extreme value copulas, is shown to be completely determined by its tail dependence functions. For a vine copula built from a set of bivariate copulas, its tail dependence function can be expressed recursively by the tail dependence and conditional tail dependence functions of lower-dimensional margins. The effect of tail dependence of bivariate linking copulas on that of a vine copula is also investigated. © 2009 Elsevier Inc. All rights reserved.

Orthant tail dependence of multivariate extreme value distributions

Volume 100, Issue 1, January 2009, Pages 243-256
Li, H.

The orthant tail dependence describes the relative deviation of upper- (or lower-) orthant tail probabilities of a random vector from similar orthant tail probabilities of a subset of its components, and can be used in the study of dependence among extreme values. Using the conditional approach, this paper examines the extremal dependence properties of multivariate extreme value distributions and their scale mixtures, and derives the explicit expressions of orthant tail dependence parameters for these distributions. Properties of the tail dependence parameters, including their relations with other extremal dependence measures used in the literature, are discussed. Various examples involving multivariate exponential, multivariate logistic distributions and copulas of Archimedean type are presented to illustrate the results. © 2008 Elsevier Inc. All rights reserved.
Likelihood ratio order of the second order statistic from independent heterogeneous exponential random variables

Volume 100, Issue 5, May 2009, Pages 952-962
Zhao, P. | Li, X. | Balakrishnan, N.

Let $X_1, \ldots, X_n$ be independent exponential random variables with respective hazard rates $\lambda_1, \ldots, \lambda_n$, and let $Y_1, \ldots, Y_n$ be independent exponential random variables with common hazard rate $\lambda$. This paper proves that $X_2 : n$, the second order statistic of $X_1, \ldots, X_n$, is larger than $Y_2 : n$, the second order statistic of $Y_1, \ldots, Y_n$, in terms of the likelihood ratio order if and only if

$$\lambda \geq \frac{1}{2n-1} \left(2 \Lambda_1 + \frac{\Lambda_3 - \Lambda_1 \Lambda_2, \Lambda_1 2 - \Lambda 2}) \right)$$

with $\Lambda_k = \sum_{i=1}^{n} \lambda_i k$, $k = 1, 2, 3$. Also, it is shown that $X_2 : n$ is smaller than $Y_2 : n$ in terms of the likelihood ratio order if and only if

$$\lambda \leq \frac{\sum_{i=1}^{n} \lambda_i - \max_{1 \leq i \leq n} \lambda_i, n - 1}{n - 1}.$$ 


A test for the mean vector with fewer observations than the dimension

Srivastava, M.S. | Du, M.

In this paper, we consider a test for the mean vector of independent and identically distributed multivariate normal random vectors where the dimension $p$ is larger than or equal to the number of observations $N$. This test is invariant under scalar transformations of each component of the random vector. Theories and simulation results show that the proposed test is superior to other two tests available in the literature. Interest in such significance test for high-dimensional data is motivated by DNA microarrays. However, the methodology is valid for any application which involves high-dimensional data. © 2006 Elsevier Inc. All rights reserved.

Change detection in autoregressive time series

Gombay, E.

Autoregressive time series models of order $p$ have $p + 2$ parameters, the mean, the variance of the white noise and the $p$ autoregressive parameters. Change in any of these over time is a sign of disturbance that is important to detect. The methods of this paper can test for change in any one of these $p + 2$ parameters separately, or in any collection of them. They are available in forms that make one-sided tests possible, furthermore, they can be used to test for a temporary change. The test statistics are based on the efficient score vector. The large sample properties of the change-point estimator are also explored. © 2007 Elsevier Inc. All rights reserved.

Robust parameter estimation with a small bias against heavy contamination

Volume 99, Issue 9, October 2008, Pages 2053-2081
Fujisawa, H. | Eguchi, S.

In this paper we consider robust parameter estimation based on a certain cross entropy and divergence. The robust estimate is defined as the minimizer of the empirically estimated cross entropy. It is shown that the robust estimate can be regarded as a kind of projection from the viewpoint of a Pythagorean relation based on the divergence. This property implies that the bias caused by outliers can become sufficiently small even in the case of heavy contamination. It is seen that the asymptotic variance of the robust estimator is naturally overweighted in proportion to the ratio of contamination. One may surmise that another form of cross entropy can present the same behavior as that discussed above. It can be proved under some conditions that no cross entropy can present the same behavior except for the cross entropy considered here and its monotone transformation. © 2008 Elsevier Inc. All rights reserved.

Learning from dependent observations

Volume 100, Issue 1, January 2009, Pages 175-194

Most Cited Journal of Multivariate Analysis Articles http://www.journals.elsevier.com/journal-of-multivariate-analysis/most-c...
Learning from dependent observations
Volume 100, Issue 1, January 2009, Pages 175-194
Steinwart, I. | Hush, D. | Scovel, C.

In most papers establishing consistency for learning algorithms it is assumed that the observations used for training are realizations of an i.i.d. process. In this paper we go far beyond this classical framework by showing that support vector machines (SVMs) only require that the data-generating process satisfies a certain law of large numbers. We then consider the learnability of SVMs for $\alpha$-mixing (not necessarily stationary) processes for both classification and regression, where for the latter we explicitly allow unbounded noise.

Applications of average and projected systems to the study of coherent systems
Volume 101, Issue 6, July 2010, Pages 1471-1482
Navarro, J. | Spizzichino, F. | Balakrishnan, N.

In this paper, we introduce the concepts of average and projected systems associated to a coherent (parent) system. We analyze several aspects of these notions and show that they can be useful tools in studying the performance of coherent systems with non-exchangeable components. We show that the average and projected systems are especially useful in studying the tail behavior of reliability, hazard rate and mean residual life functions of the parent system and also in obtaining the tail best systems (under different criteria) by permuting the components at the system structure. Moreover, they can be useful in assessing how the asymmetry of the joint distribution of the component lifetimes (with respect to permutations of the components in the system structure) affects the system performance. © 2010 Elsevier Inc. All rights reserved.

Log-linear Poisson autoregression
Volume 102, Issue 3, March 2011, Pages 563-578
Fokianos, K. | Tjøstheim, D.

We consider a log-linear model for time series of counts. This type of model provides a framework where both negative and positive association can be taken into account. In addition time dependent covariates are accommodated in a straightforward way. We study its probabilistic properties and maximum likelihood estimation. It is shown that a perturbed version of the process is geometrically ergodic, and, under some conditions, it approaches the non-perturbed version. In addition, it is proved that the maximum likelihood estimator of the vector of unknown parameters is asymptotically normal with a covariance matrix that can be consistently estimated. The results are based on minimal assumptions and can be extended to the case of log-linear regression with continuous exogenous variables. The theory is applied to aggregated financial transaction time series. In particular, we discover positive association between the number of transactions and the volatility process of a certain stock. © 2010 Elsevier Inc.

Empirical likelihood for linear models with missing responses
Volume 100, Issue 7, August 2009, Pages 1353-1366
Xue, L.

The purpose of this article is to use an empirical likelihood method to study the construction of confidence intervals and regions for the parameters of interest in linear regression models with missing response data. A class of empirical likelihood ratios for the parameters of interest are defined such that any of our class of ratios is asymptotically chi-squared. Our approach is to directly calibrate the empirical log-likelihood ratio, and does not need multiplication by an adjustment factor for the original ratio. Also, a class of estimators for the parameters of interest is constructed, and the asymptotic distributions of the proposed estimators are obtained. Our results can be used directly to construct confidence intervals and regions for the parameters of interest. A simulation study indicates that the proposed methods are comparable in terms of coverage probabilities and average lengths/areas of confidence intervals/regions. An example of a real data set is used for illustrating our methods. © 2008 Elsevier Inc. All rights reserved.

From moments of sum to moments of product
Kan, R.

We provide an identity that relates the moment of a product of random variables to the moments of different linear combinations of the random variables. Applying this identity, we obtain new formulae for the expectation of the product of normally distributed random variables and the product of quadratic forms in normally distributed random variables.
we provide an identity that relates the moment of a product of random variables to the moments of different linear combinations of the random variables. Applying this identity, we obtain new formulae for the expectation of the product of normally distributed random variables and the product of quadratic forms in normally distributed random variables. In addition, we generalize the formulae to the case of multivariate elliptically distributed random variables. Unlike existing formulae in the literature, our new formulae are extremely efficient for computational purposes. © 2007 Elsevier Inc. All rights reserved.

Order restricted inference for sequential k-out-of-n systems
Volume 99, Issue 7, August 2008, Pages 1489-1502
Balakrishnan, N. | Beutner, E. | Kamps, U.

Sequential order statistics have been introduced to model sequential k-out-of-n systems which, as an extension of k-out-of-n systems, allow the failure of some components of the system to influence the remaining ones. Based on an independent sample of vectors of sequential order statistics, the maximum likelihood estimators of the model parameters of a sequential k-out-of-n system are derived under order restrictions. Special attention is paid to the simultaneous maximum likelihood estimation of the model parameters and the distribution parameters for a flexible location-scale family. Furthermore, order restricted hypothesis tests are considered for making the decision whether the usual k-out-of-n model or the general sequential k-out-of-n model is appropriate for a given data. © 2008 Elsevier Inc. All rights reserved.

Mean residual life order of convolutions of heterogeneous exponential random variables
Volume 100, Issue 8, September 2009, Pages 1792-1801
Zhao, P. | Balakrishnan, N.

In this paper, we study convolutions of heterogeneous exponential random variables with respect to the mean residual life order. By introducing a new partial order (reciprocal majorization order), we prove that this order between two parameter vectors implies the mean residual life order between convolutions of two heterogeneous exponential samples. For the 2-dimensional case, it is shown that there exists a stronger equivalence. We discuss, in particular, the case when one convolution involves identically distributed variables, and show in this case that the mean residual life order is actually associated with the harmonic mean of parameters. Finally, we derive the "best gamma bounds" for the mean residual life function of any convolution of exponential distributions under this framework. © 2009 Elsevier Inc. All rights reserved.

On weighting of bivariate margins in pairwise likelihood
Volume 100, Issue 4, April 2009, Pages 670-685
Joe, H. | Lee, Y.

Composite and pairwise likelihood methods have recently been increasingly used. For clustered data with varying cluster sizes, we study asymptotic relative efficiencies for various weighted pairwise likelihoods, with weight being a function of cluster size. For longitudinal data, we also study weighted pairwise likelihoods with weights that can depend on lag. Good choice of weights are needed to avoid the undesirable behavior of estimators with low efficiency. Some analytic results are obtained using the multivariate normal distribution. For clustered data, a practically good choice of weight is obtained after study of relative efficiencies for an exchangeable multivariate normal model; they are different from weights that had previously been suggested. For longitudinal data, there are advantages to only include bivariate margins of adjacent or nearly adjacent pairs in the weighted pairwise likelihood. © 2008 Elsevier Inc. All rights reserved.