International Conference on Trends and Perspectives in Linear Statistical Inference

---

Book of Abstracts

---

04 – 08 July, 2022
Tomar, Portugal

Edited by

Daniel Klein
Institute of Mathematics
P. J. Šafárik University, Košice, Slovakia

and

Francisco Carvalho
Instituto Politécnico de Tomar, Tomar, Portugal

Contents

Part I

Introduction

The International Conference on Trends and Perspectives in Linear Statistical Inference, LinStat2022, will be held 4 - 8 July, 2022, in Tomar, Portugal, at the Polytechnic Institute of Tomar. This is the follow-up of the LinStat series held in Będlewo, Poland (2008, 2012, 2018, 2021), in Tomar, Portugal (2010), in Linköping, Sweden (2014), and in İstanbul, Turkey (2016).

The purpose of the conference is to bring together researchers sharing an interest in a variety of aspects of statistics and its applications and offer them a possibility to discuss current developments in these subjects. The conference will mainly focus on a number of topics. The topics that have been selected so far include estimation, prediction and testing in linear models, robustness of relevant statistical methods, estimation of variance components appearing in linear models, generalizations to nonlinear models, design and analysis of experiments, including optimality and comparison of linear experiments.

The work of young scientists is highly appreciated. The Scientific Committee will award the best presentation and best poster. The awarded will be Invited Speakers at the next edition of LinStat.

The conference will include invited talks given by

- S. Ejaz Ahmed (Canada)
- Carlos A. Coelho (Portugal)
- Stanislav Katina (Czech Republic)
- Célia Nunes (Portugal)
- Han Lin Shang (Australia)

as well as invited talks of Young Scientist Awards of Linstat2020 winners

- Efthymia Derezea (UK)
- Malwina Janiszewska (Poland)
- Denise Uwamariya (Sweden)

Organizers

- Polytechnic Institute of Tomar, Tomar, Portugal
- Universidade NOVA de Lisboa, Lisbon, Portugal
- Institute of Mathematics, Poznań University of Technology, Poland
- Institute of Mathematics, Faculty of Science, P. J. Šafárik University, Košice, Slovakia

Committees

The Scientific Committee for this Conference comprises

- João Tiago Mexia (Portugal) - Chair
- Augustyn Markiewicz (Poland) - Vice-Chair
- Anthony C. Atkinson (UK)
- Simo Puntanen (Finland)
- Dietrich von Rosen (Sweden)
- Müjgan Tez (Turkey)
- Götz Trenkler (Germany)
- Ivan Žežula (Slovakia)
- Roman Zmyślony (Poland)

The Organizing Committee comprises

- Francisco Carvalho (Portugal) - Chair
- Miguel Fonseca - Vice-Chair
- Cristina Andrade (Portugal)
- Katarzyna Filipiak (Poland)
- Daniel Klein (Slovakia)

Call for papers

We are pleased to announce special issues of *Journal of Statistical Computation and Simulation (JSCS)* devoted to LinStat2022, which will include selected papers strongly correlated to the talks of the conference.

Moreover, the participants presenting topics related to multivariate analysis are encouraged to submit original work to the *Journal of Multivariate Analysis (JMVA)*.

All papers submitted to the journals must meet the publication standards of JSCS or JMVA, respectively, and will be subject to normal refereeing procedure. Note also, that they must satisfy aim and scope of the chosen journal (JSCS or JMVA), otherwise, they will not be considered for publication.

Papers submitted to JSCS should be prepared according to the instructions found on the journal web site and the submissions should be sent to linstat2020@gmail.com email address (with a subject "LinStat2022 JSCS submission"). For each paper at least two potential referees should be indicated. After preliminary verification of the submission by guest editors the authors will be invited to submit their paper through JSCS submission system.

Part II

Program

LinStat'2022 - General Program

	MONDAY		TUESDAY	WEDNESDAY	THURSDAY	FRIDAY
8:50 - 9:00	opening					
9:00 - 9:50	S. E. Ahmed		H. L. Shang (*)	S. Katina	C. A. Coelho	D. Uwamariya (*)
	PARALLEL SESSIONS					
10:00-10:20	J. T. Mexia	T. Oliveira	V. de Oliveira	Y. Liang	C. Miranda	I. Zezula
10:20-10:40	P. Rodrigues	A. Seijas-Macias	V. Makogin	K. Filipiak	A. Rocha	R. Zmyslony
10:40-11:00	~~O. Yolcu~~	J. Pereira	M. Spyropoulou	M. John	~~I. Yaman~~	A. Markiewicz
	COFFEE BREAK					
11:30-11:50	U. Beyaztas	M. F. Teodoro	N. Loperfido	E. Moreira	D. von Rosen	S. Puntanen
11:50-12:10	M. Tez	M. Andrade	S. Mazur	M. Chen	S. Anatolyev	O. Baksalary
12:10-12:30	F. Ozkurt	A. Carrasco	C. Franceschini	S. Kim	D. Ferreira	P. Antunes
12:30-12:50	S. Kilicoglu		M. Giacalone	C. Andrade	X. Hu	E. Fiserova
						closing
13:00	LUNCH					
14:30-15:20	E. Dereza		M. Janiszewska		C. Nunes	
15:30-15:50	M. Capanu		C. Dias		M. Norouzirad	
15:50-16:10	E. Gasana		C. Santos		X. Wang	
16:10-16:30	M. Mansour		A. Catarinha		A. Touazi	
	COFFEE BREAK				COFFEE BREAK	
16:50-17:10	M. Mokrzycka		P. Monteiro		S. Kim	
17:10-17:30	D. Dai		C. Francisco		J. Park	
				14:00 Social Program		
					Conference Dinner	

Part III

Invited Speakers

Sparse estimation strategies in linear mixed effect models for high-dimensional data application

S. Ejaz Ahmed

Brock University, Ontario, Canada

Abstract

In a host of business applications, biomedical and epidemiological studies, the problem of multicollinearity among predictor variables is a frequent issue in longitudinal data analysis for linear mixed models (LMM). We consider efficient prediction strategies for high-dimensional data application.

Specifically, I discussed improved estimation of the fixed effects parameters of the LMM when it is Judiciously assumed that model is sparse. We propose the pretest and shrinkage low and high dimensional estimation strategies using the ridge full model as the base estimator. We establish the asymptotic distributional bias and risks of the suggested estimators and investigate their relative performance with respect to the ridge full model estimator in low dimensional case. Furthermore, we compare the numerical performance of the penalized estimators with the pretest and shrinkage ridge estimators. The methodology is investigated using simulation studies and then demonstrated on an application exploring how effective brain connectivity in the default mode network (DMN) may be related to genetics within the context of Alzheimer's disease.

A high-performance likelihood ratio test for high-dimensional MANOVA

Carlos A. Coelho

Nova University of Lisbon, Portugal

Abstract

Is it possible to develop a likelihood ratio test for high-dimensional MANOVA? Would such test perform well? Would it be able to outperform existing tests? Would it be applicable to extremely small samples? Would it be applicable to non-normal random variables, as uniform, extremely skewed distributions, or even heavy tailed distributions with success? Would it have a nice, rather simple to compute and well performing, asymptotic distribution? Although most researchers in the area of Statistics may think that the answer to all of the above questions is a 'No', it will be shown that it is actually a 'Yes'! And to all of them!

The presentation also shows the advantages of the LRT developed in relation to other existing tests, namely in terms of power and control of the Type I error rate. A quite simple, but very well-performing, Normal asymptotic distribution is obtained for the test statistic and a number of simulation results show the advantages of the LRT and its adequacy even for non-normal, highly skewed and/or heavy tailed distributions.

Frequency estimation of astronomical light curves with red noise

Efthymia Derezea, Alfred Kume and Dirk Froebrich

University of Kent, Canterbury, UK

Abstract

Estimating the periodic behavior of light curves, which are a type of unequally-spaced time series describing a star's brightness over time, is of great interest in Astronomy. In this work we consider a harmonic model with additive red noise. This is a type of correlated noise appearing often in astronomy and it is the equivalent of assuming an AR(1) structure for the residuals of equally-spaced data. We generalize a result from [1] and show that our frequency estimate is consistent and asymptotically normal. Furthermore, we consider a type of generalized F-test in order to decide if our estimate is true or the product of noise fluctuations. Finally, we report our simulation results and apply our methods to light curves from the Hunting Outbursting Young Stars citizen science project.

Keywords

Unequally-spaced time series, Red noise, Power correlation, Asymptotics.

References

[1] Reimann, J.D. Frequency estimation using unequally-spaced astronomical data (Doctoral dissertation, University of California, Berkeley).

[2] Ibragimov, I.A., Has'Minskii, R.Z (2013). *Statistical estimation: asymptotic theory*. Springer.

Estimation of linearly structured block covariance matrix

Malwina Janiszewska[1], Augustyn Markiewicz[1] and Monika Mokrzycka[2]

[1] Poznań University of Life Sciences, Poland
[2] Polish Academy of Sciences, Poznań, Poland

Abstract

In this paper, the study of the relation between two groups of characteristics observed on sample units is of interest. The observation received from the experimental unit has a form of vector and it can be divided into two subvectors, corresponding with two group of characteristics. The relation between features is expressed in cross-covariance matrix, which structure is considered. The covariance structure with symmetric positive definite diagonal blocks and linearly structured off-diagonal blocks is assumed. Moreover, the covariance matrix with all blocks linearly structured is studied. The aim of the research is to determine structured estimator of the covariance matrix with good statistical properties (such as consistency), which is both (i) positive definite and (ii) well-conditioned. In the case of structured covariance matrix the maximum likelihood estimation procedure can provide to complex numerical problem. Alternative approach is proposed, which is based on least squares method [3], shrinkage method [1], and additional algebraic improvement. New approach, called quasi shrinkage estimation, is much faster and it allows to determine estimators in simple form, with good statistical properties. The obtained estimator has a similar form as the estimators given in [2]. Simulations study are performed to compare maximum likelihood estimators and quasi shrinkage estimators of structured covariance matrices by some criteria. Furthermore, characteristics of quadratic spaces of block matrices are derived.

Keywords

Block covariance matrix, Least squares method, Linear structure, Quadratic space, Shrinkage method.

References

[1] Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88(2)*, 365–411.
[2] Markiewicz, A. and Mieldzioc, A. (2022). Improved estimators of linearly structured covariance matrices. Submitted.
[3] Ohlson, M. and von Rosen, D. (2010). Explicit estimators of parameters in the Growth Curve model with linearly structured covariance matrices. *Journal of Multivariate Analysis 101*, 1284–1295.

Functional data analysis of three-dimensional surface data

Stanislav Katina

[1] Masaryk University, Brno, Czech Republic
[2] Czech Academy of Sciences, Prague, Czech Republic

Abstract

The advent of high-resolution imaging has made surface shape data widespread. Methods for the analysis of shapes based on points (landmarks) are well established, but high-resolution data require a functional approach.
First, a systematic and consistent description of each surface shape (using landmarks, curves (semi-landmarks), and surface patches (semi-landmarks)) and a method of automatic identification of this using penalised regression models with constraints and conditions are described. Second, the registration of curves and surfaces in functional form is discussed. Then the functional principal component (PC) analysis of curves and surfaces and PC subspaces where interesting behaviour, such as population differences, is exhibited (rather than on individual PCs), are presented. Finally, functional regression models of curves and surfaces are defined.
All these ideas are developed and illustrated in the important context of the human facial shape of healthy individuals, patients before and after orthognathic surgery, or patients with psychotic or other disorders and controls, with a strong emphasis on effective visual communication of effects of interest. All the methods presented here are implemented in R as part of the development of the face3d package.

Keywords

Curves, Surfaces, Automatic identification, Penalised regression models, Functional registration, Functional principal component analysis, Functional regression, Human face.

References

[1] Bowman, A.W., Katina, S., Smith, J. and Brown, D. (2015). Anatomical curve identification. *Computational Statistics and Data Analysis 86(6)*, 52–64.
[2] Prasad, S., Katina, S., Hennessy, R.J., Murphy, K.C., Bowman, A.W. and Waddington, J.L. (2015). Craniofacial dysmorphology in 22q11.2 deletion syndrome by 3D laser surface imaging and geometric morphometrics: illuminating the developmental relationship to risk for psychosis. *American Journal of Medical Genetics Part A 167(3)*, 529–536.
[3] Katina, S., McNeil, K., Ayoub, A., Guilfoyle, B., Khambay, B., Siebert, P., Sukno, F., Rojas, M., Vittert, L., Waddington, J.L, Whelan, P.F, and Bowman, A.W. (2016). The definitions of three-dimensional landmarks on the human face: an interdisciplinary view. *Journal of Anatomy 228(3)*, 355–365.
[4] Vittert, L., Katina, S., Ayoub, A., Khambay, B. and Bowman, A.W. (2018). Assessing the outcome of orthognathic surgery by three-dimensional soft tissue analysis. *International Journal of Oral and Maxillofacial Surgery, 47(12)*, 1587–1595.

[5] Katina, S., Kelly, B.D., Rojas, M.A., Sukno, F.M., McDermott, A., Hennessy, R.J., Lane, A., Whelan, P.F., Bowman, A.W. and Waddington, J.L. (2020). Refining the resolution of craniofacial dysmorphology in bipolar disorder as an index of brain dysmorphogenesis. *Psychiatry research 291*, 113243.

[6] Vittert, L., Bowman, A.W. and Katina, S. (2020). A hierarchical curve-based approach to the analysis of manifold data. *Annals of Applied Statistics 13(4)*, 2539–2563.

[7] Katina, S., Vittert, L. and Bowman, A.W. (2021). Data Analysis and Visualisation of Three-dimensional Surface Shape. *Journal of Royal Statistical Society, Series C 70(3)*, 691–713.

ANOVA with random sample sizes: an overview and applications

Célia Nunes

Universiddade da Beira Interior, Covilhã, Portugal

Abstract

Analysis of variance (ANOVA) is one of the most frequently used statistical analyses in practical applications. It is routinely used in several research areas, namely in Medical and Biomedical Sciences, Agriculture, Social Sciences, to name just a few. Despite being widely used on the assumption that sample dimensions are known, there are many relevant situations in which these dimensions are not known in advance. Such situations frequently occur when a fixed time period is established for collecting the observations. In this presentation we show that this may be overcome when we carry out ANOVA assuming the sample sizes as realizations of independent random variables. This approach must be based on an adequate choice of the distribution of these random variables. There are two families of such distributions, according to the existence or non-existence of an upper bound for the sample sizes. Examples with distributions of these two families will be considered, in which their choice is based on practical situations. The applicability of the proposed approach is illustrated through some applications, considering real and simulated data. With this presentation, we intend to give an overview of our latest advances in the field of "ANOVA with random sample sizes", considering fixed and mixed effects models.

Keywords

Fixed effects models, Mixed models, F-tests, Random sample sizes, Applications to real data, Simulation studies.

Acknowledgements

This work was partially supported by national founds of FCT-Foundation for Science and Technology under UIDB/00212/2020.

References

[1] Nunes, C., Ferreira, D., Ferreira, S.S. and Mexia, J.T. (2014). Fixed effects ANOVA: An extension to samples with random size. *Journal of Statistical Computation and Simulation 84(11)*, 2316–2328.

[2] Nunes, C., Capistrano, G., Ferreira, D., Ferreira, S.S. and Mexia, J. T. (2019). Exact critical values for one-way fixed effects models with random sample sizes. *Journal of Computational and Applied Mathematics 354*, 112–122.

[3] Nunes, C., Mário, A., Ferreira, D., Ferreira, S.S. and Mexia, J.T. (2019). Random sample sizes in orthogonal mixed models with stability. *Computational and Mathematics Methods 1(5)*.

[4] Nunes, C., Mário, A., Ferreira, D., Moreira, E., Ferreira, S.S. and Mexia, J.T. (2022). An algorithm for simulation in mixed models with crossed factors considering the sample sizes as random. *Journal of Computational and Applied Mathematics 404*, 113463.

Functional capital asset pricing model

Ufuk Beyaztas[1], Kaiying Ji[2], Han Lin Shang[3] and Eliza Wu[2]

[1] Marmara University, Istanbul, Turkey
[2] University of Sydney, Australia
[3] Macquarie University, Australia

Abstract

The capital asset pricing model (CAPM) is a well-celebrated model to capture a linear relationship between the daily returns of an asset and a market index. We extend this model to the high-frequency setting by proposing a functional CAPM. The functional CAPM is a stylized example of function-on-function linear regression. The bivariate functional regression coefficient measures a linear relationship between intraday cumulative returns of an asset and a market index. Using the Dow-Jones Industrial Average index and its constituent stocks, we investigate the goodness-of-fit of the functional CAPM for the intraday return of an asset.

Keywords

Cumulative intraday returns, Dow-Jones Industrial Average index, Function-on-function linear regression, Regression coefficient function.

Large deviations of extremal eigenvalues of sample covariance matrices

Denise Uwamariya and Xiangfeng Yang

Linköping University, Sweden

Abstract

Large deviations of the largest and smallest eigenvalues of $\mathbf{X}\mathbf{X}^\top/n$ are studied in this note, where $\mathbf{X}_{p\times n}$ is a $p \times n$ random matrix with independent and identically distributed (i.i.d.) sub-Gaussian entries. The assumption imposed on the dimension size p and the sample size n is $p = p(n) \to \infty 1$ with $p(n) = o(n)$. This study generalizes one result obtained in [?].

Keywords

Large deviations, Sample covariance matrices, Extremal eigenvalues.

References

[1] Bai, Z.D. and Yin, Y.Q. (2008). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *In Advances In Statistics. World Scientific*, 108–127.

[2] Fey, A., van der Hofstad, R. and Klok, M.J. (2008). Large deviations for eigenvalues of sample covariance matrices, with applications to mobile communication systems. *Advances in Applied Probability 40(4)*, 1048–1071.

[3] Geman, S. (1980). A limit theorem for the norm of random matrices. *The Annals of Probability*, 252–261.

[4] Jonsson, D. (1983). On the largest eigenvalue of a sample covariance matrix. Uppsala University, Department of Mathematics.

[5] Rogers, C. (1963). Covering a sphere with spheres. *Mathematika 10(2)*, 157–164.

[6] Singull, M., Uwamariya, D. and Yang, X. (2021). Large-deviation asymptotics of condition numbers of random matrices. *Journal of Applied Probability 58(4)*, 1114–1130.

[7] Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv:1011.3027*.

[8] Yin, Y.Q., Bai, Z.D. and Krishnaiah, P.R. (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability theory and related fields 78(4)*, 509–521.

Part IV

Abstracts

Leave-cluster-out and variance estimation

Stanislav Anatolyev

CERGE-EI, Charles University, Prague, Czech Republic

Abstract

We introduce the leave-cluster-out (LCO) machinery for clustered samples, a generalization of leave-one-out methods that prove useful for independent data. We use LCO to construct an estimator of the asymptotic variance of the OLS estimator in a linear regression characterized by possibly numerous regressors and arbitrary within-cluster heteroskedasticity. We show consistency of the LCO variance estimator when regressors may be many, regression errors may be heteroskedastic, clusters may be unbalanced and heterogeneous, and cluster sizes may be moderately large. Simulations reveal amazing robustness of the LCO estimator to regressor numerosity and heteroskedasticity.

Keywords

Linear regression, Ordinary least squares, Heteroskedasticity, Many regressors, Leave-out estimation, Variance estimation.

A statistical viewpoint regarding the Fire Weather Index (FWI) for Portugal

Cristina Andrade

[1] Instituto Politécnico de Tomar, Portugal
[2] Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal

Abstract

Portugal (PT) endures extreme forest fires, with a record of burned area in 2017 (ICNF, 20212). The relevance of meteorological conditions for the incidence of conditions prone for the occurrence of forest fires is well known, hence the ability to anticipate their impact on daily fire occurrence and related behavior is one of the major goals of researchers in this area of expertise. In Portugal, the Canadian Forest Service Fire Weather Index (FWI) System is the most used danger index [2, 3], and its forecast is given by the Instituto Português do Mar e da Atmosfera (IPMA). This index is a combination of three fuel moisture codes and two related with fire intensity potential, thus providing a danger risk index. This study aims at analyzing if there is a stable order of correlations of the six indexes with FWI, during several time-periods comprising the 2017 major wildfires in PT. For each time-period the ranks of the z values were attained by applying the Fisher's z-transformation [4] to adjust correlation coefficients and stabilize the variance, and the equality of the ordering was tested using Friedman's test [5]. It may be concluded that the use of a sampling scheme based on prime basis factorial enabled us to get a balanced coverage for every time-period, avoiding clustering on latitudes, longitudes or on time. Thus, we can consider our results as balanced and covering the dimensions of latitude, longitude, and time.

References

[1] Instituto de Conservação da Natureza e Florestas - Sistema de Gestão de Informação de Incêndios Florestais (ICNF) (accessed on 31 november 2021).
[2] van Wagner, C.E. and Pickett, T.L. (1985). Equations and FORTRAN program for the Canadian Forest Fire Weather Index System. Forestry Technical Report 33. Canadian Forestry Service, Petawawa National Forestry Institute, Chalk River, Ontario, Canada, pp. 18.
[3] Lawson, B.D. and Armitage, O.B. (2008). Weather Guide for the Canadian Forest Fire Danger Rating System. Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre, Edmonton, Alberta, pp. 73.
[4] Kendal, M.G. and Stuart, A. (1961). *The advance theory of statistics, Vol. 2.* Inference and Relationship, Charles Griffin and Company, Lda., London.
[5] Conover, W.J. (1999). *Practical Nonparametric Statistics, 3rd Edition.* John Wiley & Sons.

Estimation of variance components in the mixed-Effects models: the case in which the random part follows a Gamma distribution

Patrícia Antunes[1], Sandra S. Ferreira[1], Dário Ferreira[1] and João T. Mexia[2]

[1] University of Beira Interior, Covilhã, Portugal
[2] Nova University of Lisbon, Portugal

Abstract

In our talk we present the mixed-effects linear model in which the random part follows a Gamma distribution. The mixed-effects linear model has received considerable attention from both theoretical and practical points of view due to its extensive applications. We consider the estimators of the cumulants of these models as linear combinations of first order cumulants with suitably chosen independent coefficients from a random part. Moreover, we simulate the variance components and show that their estimates can achieve good performance.

Keywords

Cumulants, Mixed-effects model, Method of Moments, Gamma distribution.

References

[1] Demidenko, E. (2004). *Mixed Models: Theory and Applications.* John Wiley & Sons.
[2] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions,* 2^{nd} *edition.* John Wiley & Sons.
[3] Koutrouvelis, I.A., Canavos, G.C. and Kallioras, A.G. (2010). Cumulants Plots for Assessing the Gamma Distribution. *Communications in Statistics-Theory and Methods 39(4)*, 626–641.
[4] Lehmann, E.L. (1986). *Testing Statistical Hypotheses,* 2^{nd} *edition.* John Wiley & Sons.
[5] Zheng, Q. (2002). Computing relations between moments and cumulants. *Computational Statistics 17(4)*, 507–515.

ANOVA and the minimal least squares estimator

Oskar Maria Baksalary[1] and Götz Trenkler[2]

[1] Adam Mickiewicz University, Poznań, Poland
[2] Dortmund University of Technology, Germany

Abstract

The one-way classification model of ANOVA (analysis of variance) is considered by exploiting an original approach based on an expression for the Moore-Penrose inverse of a columnwise partitioned matrix. In consequence, various original characteristics of the model are established. It is shown, inter alia, that the common estimator of ANOVA, obtained as a solution to the normal equations, is not necessarily the minimal least squares estimator.

Keywords

Analysis of variance, Estimation theory, Expectation value, Experimental data processing, Moore-Penrose inverse, Partitioned matrix, UMVU estimator.

Machine learning-based functional time series forecasting: Application to age-specific mortality rates

Ufuk Beyaztas[1] and Han Lin Shang[2]

[1] Marmara University, Istanbul, Turkey
[2] Macquarie University, Sydney, Australia

Abstract

We propose a functional time series method to obtain accurate multistep-ahead forecasts for age-specific mortality rates. The dynamic functional principal component analysis method is used to decompose the mortality curves into dynamic functional principal components and their associated principal component scores. Machine learning-based multi-step-ahead forecasting strategies, which automatically learn the underlying structure of the data, are used to obtain the future realization of the principal component scores. The forecasted mortality curves are obtained by combining the dynamic functional principal components and forecasted principal component scores. The point and interval forecast accuracy of the proposed method is evaluated using six age-specific mortality datasets and compared favorably with four existing functional time series methods.

Keywords

Direct prediction strategy, Dynamic functional principal component analysis, Long-run covariance, Machine learning, Recursive prediction strategy.

Acknowledgements

This work was supported by The Scientific and Technological Research Council of Turkey (TUBITAK) (grant no: 120F270).

Multivariate collective risk models - inference and special case

Ana Cantarinha[1], Elsa Moreira[2] and João T. Mexia[2]

[1] Instituto Politécnico de Beja, Portugal
[2] Universidade Nova de Lisboa, Portugal

Abstract

Univariate collective models have played an important role in Actuarial Mathematics. The inference about these models is usually made for the totals of claims.
We now present a multivariate version of these models that may be of interest, as a special case, for application in forest fires. The inference in this case is now made for the total burnt area and the number of fires.

Keywords

Collective models, Asymptotic distributions, Confidence intervals, Risk theory.

References

[1] Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A. and Nesbet, C.J. (1986). *Actuarial Mathematics*. The Society of Actuaries.
[2] Cantarinha, A., Moreira, E., Oliveira, M. and Mexia, J.T. (2022). A risk model for forest fires based on asymptotic results for multivariate collective models. Single models ans structured families of models. *Communications in Statistics - Theory and Methods*.
[3] Mexia, J.T. (1990). *Variance free models*, Trabalhos de Investigação, No Departamento de Matemática, Faculdade de Ciências e Tecnologia. Universidade Nova de Lisboa.
[4] Marques, S., Borges, J.G., Garcia-Gonzalo, J., Moreira, F., Carreiras, J.M.B., Oliveira, M.M., Cantarinha, A., Botequim, B. and Pereira, J.M.C. (2011). Characterization of wildfires in Portugal. *European Journal of Forest Research 130(5)*, 775–784.

Subsampling based variable selection for generalized linear models

Marinela Capanu[1], Mihai Giurcanu[2], Colin B. Begg[1] and Mithat Gönen[1]

[1] Memorial Sloan Kettering Cancer Center, New York, USA
[2] University of Chicago, Illinois, USA

Abstract

In earlier work we introduced a novel variable selection method for low dimensional linear models involving repeatedly splitting the data to establish an optimal variable selection cutoff. Building on this approach, in this article we adapt our strategy for the generalized linear model setting. We propose repeatedly subsampling the data, minimizing the Akaike's Information Criterion (AIC) over a sequence of nested models for each subsample, and including in the final model those predictors selected in the minimum AIC model in a large fraction of the subsamples. We name this novel approach AIC OPTimization via Subsampling (OPTS-AIC). We also introduce new techniques which involve optimization of the screening threshold over repeated subsamples. In an extensive simulation study examining a variety of proposed variable selection methods we show that, although no single method uniformly outperforms the others in all the scenarios considered, OPTS-AIC enjoys superior performance compared to candidate methods in many settings. We illustrate the methods by applying them to logistic and Poisson regressions and discuss extensions to the high-dimensional setting.

Keywords

AIC, Regression, Screening threshold, Subsampling, Variable selection.

References

[1] Capanu, M., Giurcanu, M., Begg, C.B. and Gönen, M. (2020). Optimized variable selection via repeated data splitting. *Statistics in Medicine 39*, 2167–2184.

Analyzing the impact of different countries' approaches to the COVID-19 pandemic on their cumulative infection curves by using nonparametric density regression and clustering methods

Mason Chen

Stanford OHS, USA

Abstract

This project aims to study patterns in different national responses to the pandemic and their effect on controlling the outbreak through a nonparametric density and regression technique and various clustering tools. Seven different countries, United States of America, Italy, South Korea, Taiwan, United Kingdom, Brazil, and India were chosen for analysis due to their varied methods and successes in handling COVID-19 infection. First, a novel nonparametric density technique was utilized to accurately and consistently partition each country's COVID-19 cumulative growth curve into different waves and phases, and each phase was modeled using linear, quadratic, or logarithmic regression. Every country's model variables, which are the slope, r-square value, duration, and model type of each phase, were then connected to real-life factors, such as cluster outbreaks, government regulations, and the availability of healthcare resources. Multivariate correlation was conducted to uncover the relationships between model variables, and running the variable clustering algorithm showcased which model variable from a previous phase would be a good predictor for the infection situation in the following phase. Finally, based on the multivariate correlation and variable clustering results, the most important variables were used in hierarchical clustering to identify and explain the most similar and most different countries.

Keywords

COVID-19, Phase modeling, Nonparametric density, Regression, Cluster variables, Multivariate correlation, Hierarchical clustering.

Estimation of kronecker structured covariance based on modified Cholesky decomposition

Deliang Dai

Linnaeus University, Växjö, Sweden

Abstract

This paper is to investigate covariance estimation problems for highdimensional matrix-valued data. We propose a covariance estimator for the matrix-valued data from penalized matrix normal likelihood. Modified Cholesky decomposition of the covariance matrix is utilized to construct positive definite estimators. The method is applied for identify parsimony and for producing a statistically efficient estimator of a large covariance matrix of matrix-valued data. The consistent property of the proposed estimator is proven. Simulation results and a real data example are illustrated.

Keywords

Multivariate longitudinal data, Covariance matrix estimation, Modified Cholesky decomposition, Shrinkage penalty.

Comparing the production stability of common wheat genotype

Cristina Dias[1,3], Carla Santos[2,3] and João T. Mexia[3]

[1] Instituto Politécnico de Portalegre, Portugal
[2] Instituto Politécnico de Beja, Portugal
[3] Universidade Nova de Lisboa, Portugal

Abstract

In plant breeding, the aim is to obtain varieties that can be used in as wide areas as possible. Thus, cultivar comparison trials (cultivated varieties) tend to be integrated into networks. Given their robustness, the individual trials of these networks are often of the randomized block type. The new varieties of common wheat have a high productive potential, high technological quality, and even resistance to the main diseases and pests, which may create value in the grain sector in Portugal. Before selection, the new genotypes are evaluated in the Regional Trial in different environments (location combinations, period crop, and year). The variation of the environmental conditions makes the classification of genotypes different from one environment to another, that is, it is assumed that the interaction is significant. Different genotype responses to environmental variation make it difficult to identify the desired genotypes. Linear regression (LR) analysis has been frequently used to assess the stability of genotypes. The AMMI model is an alternative statistical procedure and is widely used. Given the relevance of these two models, it is important to compare them to determine which is the most suitable to select wheat genotypes with high and stable production. This work aims to compare the values of the GE interaction obtained using the AMMI model with those obtained through LR and, as proposed by Finlay and Wilkinson (1963), to compare the production stability of common durum wheat genotypes. Our results showed that The AMMI model is more efficient in describing the GE interaction of wheat genotypes evaluated in different environments than the analysis by LR. The CELTA cultivar is unstable but has the highest production; TE9206 is moderately stable and has high production; HELVIO and TROVADOR are the most stable, with production above average TE9008 and TE9204 are the most unstable, with production below average, TE9110 is unstable and has low production.

Keywords

AMMI model, Environmental conditions, Linear regression, Stability of genotypes.

Acknowledgments

This work is funded by national funds through the FCT - Fundaçãno para a Ciência e a Tecnologia, I.P., under the scope of the project UIDB/00297/2020 (Center for Mathematics and Applications).

References

[1] Pereira, D.G. and Mexia, J.T. (2003). Reproducibility of joint regression analysis. *Colloquium Biometryczne 33*, 279–293.

[2] van Eeuwijk, F.A. (1995). Linear and bilinear models for the analysis of multi-environment trials: An inventory of models. *Euphytica 33*, 279–293.
[3] Gauch, H. and Piepho, P.P. (2008). Statistical Analysis of Yield Trials by AMMI and GGE: Further Considerations. *Crop Science 48*, 866–889.
[4] Scheffé, H. (1959). *The Analysis of Variance*. John Wiley & Sons.

Multi-treatment regression designs in mixed models

Dário Ferreira[1], Sandra S. Ferreira[1], Célia Nunes[1] and João T. Mexia[2]

[1] University of Beira Interior, Covilhã, Portugal
[2] Nova University of Lisbon, Portugal

Abstract

We consider functions of mixed models, with the same variance-covariance matrix and mean vectors associated to a base design. We then study the action of the factors in that designs on the regression coefficients, thus having a multi-treatment regression design and carry out inference on the variance components and the estimable vectors.

Keywords

Linear mixed models, Inference, Regression coefficients.

Acknowledgments
This work was partially supported by the Portuguese Foundation for Science and Technology through the projects UIDB/00212/2020 and UIDB/00297/2020.

References

[1] Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.
[2] Brown, H. and Prescott, R. (1999). *Applied Mixed Models in Medicine.* Wiley, New York.
[3] Demidenko, E. (2013). *Mixed Models: Theory and Applications with R.* Wiley, New York.
[4] Fonseca, M., Mexia, J.T. and Zmyslony, R. (2003). Estimators and tests for variance components in cross nested orthogonal models. *Discussiones Mathematicae - Probability and Statistics 23*, 175–201.
[5] Khuri, A.I., Matthew, T. and Sinha, B.K. (1998). *Statistical Tests for Mixed Linear Models.* Wiley, New York.
[6] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses.* Springer, New York.
[7] Mexia, J.T. (1987). Multi-treatment regression designs. *Trabalhos de Investigação, 1.* Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

Rao score test for covariance structures included in quadratic subspace

Katarzyna Filipiak[1], Mateusz John[1] and Yuli Liang[2]

[1] Poznań University of Technology, Poland
[2] Linnaeus University, Växjö, Sweden

Abstract

In this talk we present general hypothesis related to covariance structures included in (commutative) quadratic subspace of symmetric positive definite matrices. We derive the Rao score test statistic (RST) and we verify its con- vergence to the asymptotic chi-square distribution. For real-data example we compare the decisions about rejection of specific hypotheses using empirical distribution of RST, limiting distribution as well as the decisions taken with the use of exact likelihood ratio test statistic distribution.

Keywords

Rao score test, Covariance structure, (Commutative) quadratic subspace of symmetric positive definite matrices.

Testing the effect of covariates in a spatial functional model in the presence of heteroscedasticity

Eva Fišerová and Veronika Římalová

Palacký University, Olomouc, Czech Republic

Abstract

The aim of the contribution is to introduce a permutation-based test for the effect of covariates for a functional regression model with the heterogeneous spatial structure. In this context, a permutation of residuals from the functional regression model instead of the observations themselves is proposed. A weighted least squares model is fitted to the observations, resulting into approximately exchangeable, and thus permutable, residuals. A simulation study shows that the proposed testing procedure outperform the competitor approaches that neglect the spatial structure, both in terms of power and size. The methodology will be demonstrated on a real-world geochemical data set. The spatiotemporal models are used to analyse and reveal differences in the geochemical properties of the soil at the border between the forest and field.

Keywords

Spatiotemporal model, Functional regression, Permutation test, Heteroskedasticity.

References

[1] Ramsay, J.O. and Silverman, B.W. (2013). *Functional data analysis.* New York, Springer.

[2] Římalová, V., Menafoglio, A., Pini, A., Pechanec, V. and Fišerová, E. (2020). A permutation approach to the analysis of spatiotemporal geochemical data in the presence of heteroscedasticity. *Environmetrics 31*: e2611.

Multivariate kurtosis with the R Package MultiKurt

Cinzia Franceschini[1] and Nicola Loperfido[2]

[1] Università degli Studi di Scienze Gastronomiche di Pollenzo, Italy
[2] Università degli Studi di Urbino "Carlo Bo", Urbino, Italy

Abstract

MultiKurt is an R package purported to describe, testing and visualize multivariate kurtosis. Firstly, it computes the main kurtosis-related matrices, as the fourth-moment matrix and the cokurtosis. Secondly, it incorporates state-of-the-art algorithms for computing linear projections which either maximize, minimize, or remove kurtosis. Thirdly, MultiKurt computes Mardia's and Koziol's kurtoses, together with the related normality tests. The usage of MultiKurt is illustrated with the Iris data set.

Keywords

Kurtosis, Multivariate normality testing, Projection pursuit, R package.

References

[1] Franceschini, C. and Loperfido, N. (2020). Kurt: Performs Kurtosis-Based Statistical Analysis. R package version 1.1, https://CRAN.R-project.org/package=Kurt.
[2] Koziol, J. (1987). An alternative formulation of Neyman's smooth goodness of fit tests under composite alternatives, *Metrika 34*, 17–24.
[3] Koziol, J. (1989). A note on measures of multivariate kurtosis, *Biom. J. 31*, 619–624.
[4] Loperfido, N. (2022). Kurtosis Removal for Data Pre-Processing. *Advances in Data Analysis and Classification.* doi:10.1007/s11634-022-00498-3.
[5] Malkovich, J.F. and Afifi, A.A. (1973). On Tests for Multivariate Normality. *J. Amer. Statist. Ass. 68*, 176–179.
[6] Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika 57*, 519–530.
[7] Mòri, T.F., Rohatgi, V.K. and Székely, G.J. (1993). On multivariate skewness and kurtosis. *Theory Probab. Appl. 38*, 547–551.

Multi-alphabetic hypercubes

Carla Francisco[1], João T. Mexia[2], M. Manuela Oliveira[1] and Francisco Carvalho[2,3]

[1] Universidade de Évora, Portugal
[2] Universidade Nova de Lisboa, Portugal
[3] Instituto Politécnico de Tomar, Portugal

Abstract

These hypercubes generalize the well-known Graeco-Latin squares. In one dimension m hypercube we have p^m 'points' with m coordinates taking the values $0, 1, \ldots, p-1$. In each point we locate letters from m distinct alphabets. When we vary one of the coordinates, we get all the letters of each of the alphabets. We show, how to use vector spaces over Galois Fields to obtain such hypercubes. Moreover, these hypercubes will constitute families of pairwise orthogonal ones. This opens interesting possibilities in randomized systematic sampling for continuous media.

Keywords

Graeco-Latin squares, Hypercubes, Multi-alphabetic hypercubes.

References

[1] Antunes, P., Ferreira, S.S., Ferreira, D. and Mexia, J.T. (2021). Multiple additive models. *Communications in Statistics - Theory and Methods 50(19)*, 4649–4655.
[2] Cox, D.A. (2004). *Galois Theory.* John Wiley & Sons.
[3] Dey, A. and Mukerjee, R. (1999). *Fractional Factorial Plans.* Wiley-interscience.
[4] Hinkelmann, K. and Kempthorne, O. (2005). *Design and Analysis of Experiments: Advanced Experimental Design, Vol. 2.* Wiley.
[5] Mukerjee, R. and Wu, C.F.J. (2006). *A Modern Theory of Factorial Design.* Springer Series in Statistics.

Decentralized predictive symbiotic corporate compliance mining

Emelyne Umunoza Gasana[1,2], Dietrich von Rosen[1,3] and Martin Singull[1]

[1] Linköping University, Sweden
[2] University of Rwanda, Rwanda
[3] Swedish University of Agricultural Sciences, Uppsala, Sweden

Abstract

The exact distribution of a classification function is often complicated to allow for easy numerical calculations of misclassification errors. The use of expansions is one way of dealing with this difficulty. In this paper, approximate probabilities of misclassification of the maximum likelihood based discriminant function are established via an Edgeworth-type expansion based on the standard normal distribution for discriminating between two multivariate normal populations.

Keywords

Classification rule, Discriminant analysis, Edgeworth-type expansion, Misclassification errors.

References

[1] Anderson, T.W. (1951). Classification by multivariate analysis. *Psychometrika 16*, 31–50.
[2] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis, 3rd ed.* John Wiley & Sons.
[3] Barnard, M.M. (1935). The secular variations of skull characters in four series of Egyptian skulls. *Annals of Eugenics 6*, 352–371.
[4] Cornish, E.A. and Fisher, R. A. (1938). Moments and cumulants in the specification of distributions. *Revue de lâĂŹInstitut International de Statistique*, 307–320.
[5] Davis, A. (1976). Statistical distributions in univariate and multivariate edgeworth populations. *Biometrika 63*, 661–670.
[6] Edgeworth, F. (1905). The law of error. *Transactions of the Cambridge Philosophical Society 20*, 36–66; 113–141.
[7] Edgeworth, F. (1907). On the representation of statistical frequency by a series. *Journal of the Royal Statistical Society 70(1)*, 102–106.
[8] Fisher, R.A. (1936). The use of multiple measurements in 10 taxonomic problems. *Annals of Eugenics 7*, 179–188.
[9] Fisher, R.A. (1938). The statistical utilization of multiple measurements. *Annals of Eugenics 8*, 376–386.
[10] Fujikoshi, Y., Ulyanov, V.V. and Shimizu, R. (2011). *Multivariate Statistics: High-dimensional and Large-Sample Approximations.* Volume 760. John Wiley & Sons.
[11] Gram, J.P. (1879). *Om Rækkeudviklinger, bestemte ved Hjælp af de mindste Kvadraters Methode.* PhD thesis, Andr. Fred. Høst & Son, København.

[12] Gupta, S.S. and Panchapakesan, S. (1982). Edgeworth expansions in statistics: A brief review. *Technical report, Purdue university Lafayette in department of statistics.*

[13] Hald, A. (2000). The early history of the cumulants and the gramcharlier series. *International Statistical Review 68(2)*, 137–153.

[14] Hald, A. and Steffensen, J. (2002). *On the History of Series Expansions of Frequency Functions and Sampling Distributions*, 1873-1944. Det Kongelige Danske Videnskabernes Selskab, Copenhagen.

[15] Kollo, T. and von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices.* Volume 579. Springer Science & Business Media, Dordrecht.

[16] Kudo, A. (1959). The classificatory problem viewed as a two-decision problem. *Memoirs of the Faculty of Science, Kyushu University. Series A, Mathematics 13*, 96–125.

[17] Kudo, A. (1960). The classifactory problem viewed as a a two-decision problem-II. *Memoirs of the Faculty of Science, Kyushu University. Series A, Mathematics 14*, 63–83.

[18] Laplace, P.S. (1811). Mémoire sur les intégrales définies et leur application aux probabilités, et spécialement à la recherche du milieu quâĂŹil faut choisir entre les résultats des observations. *Mémoires de lâĂŹAcadêmie Royale des Sciences de Paris*, 279–347.

[19] Lauritzen, S.L., Thiele, T.N., Thiele, J. and Hald, A. (2002). *Thiele: Pioneer in Statistics.* Clarendon Press.

[20] Mahalanobis, P.C. (1925). Analysis of race-mixture in Bengal. *Journal and Proceedings of Asiatic Society of Bengal New series 23*, 301–333.

[21] Memon, A.Z. and Okamoto, M. (1971). Asymptotic expansion of the distribution of the Z statistic in discriminant analysis. *Journal of Multivariate Analysis 1(3)*, 294–307.

[22] Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function. *The Annals of Mathematical Statistics 34*, 1286–1301.

[23] Pearson, K. (1915). On the problem of sexing osteometric material. *Biometrika 10(4)*, 479–487.

[24] Ranga Rao, R. (1960). *Some Problems in Probability Theory.* PhD thesis, Calcutta University.

[25] Siotani, M. (1977). *Asymptotic Expansions for Error Rates and Comparison of the W-procedure and the Z-procedure in Discriminant Analysis.* North-Holland, Amsterdam.

[26] Siotani, M. and Wang, R.H. (1975). *Further Expansion Formulae for Error Rates and Comparison of the W-and Z-procedures in Discriminant Analysis.* Technical Report No. 33, Department of Statistics, Kansas State University, Manhattan, Kansas.

[27] Sitgreaves, R. (1952). On the distribution of two random matrices used in classification procedures. *The Annals of Mathematical Statistics 23*, 263–270.

[28] Sitgreaves, R. (1961). Some results on the distribution of the W-classification statistic. *Chapter 19 in Solomon, H. (ed.), Studies in Item Analysis and Prediction*, 241–261.

[29] Srivastava, M.S. and Khatri, C. (1979). *An Introduction to Multivariate Statistics.* Elsevier North-Holland, New York.

[30] Thiele, T.N. (1873). Om en tilnærmelsesformel. *Tidsskrift for mathematik 3*, 22–31.

[31] Thiele, T.N. (1889). *Forlaesinger over Almindelig Iagttagelseslaere: Sandsynlighedsregning og Mindste Kvadraters Methode.* København: Reitzel.

[32] Wald, A. (1944). On a statistical problem arising in the classification of an individual into one of two groups. *The Annals of Mathematical Statistics 15*, 145–162.

Decentralized predictive symbiotic corporate compliance mining

Massimiliano Giacalone[1], Vito Santarcangelo[2] and Diego Carmine Sinitó[2,3]

[1] University of Naples "Federico II", Naples, Italy
[2] Informatica, Trapani, Italy
[3] University of Catania, Italy

Abstract

The paper shows an innovative approach called predictive symbiotic corporate compliance mining, that is an evolution of process mining that not consider the log of a single company but a symbiotic mining of a corporate process, considering the historical and actual log of events of some companies that work in a field/business in a symbiotic approach, taking into account the national and regulatory context of reference. This evaluation consider the impact relation of each company united to a semantic mining related to the regulatory aspects using the web3 technologies. The module of process mining is implemented with a decentralized infrastructure (DApp) to collect and certify the log to reveals critical issues of the group. This decentralized approach is also applied for the development of a semantic infrastructure used to analyse the performance related to the environmental context and the compliance related to normative regulatory aspect and interpration, extracted from OSINT Open Data and recorded from SOCMINT data, thanks to a proof of reputation consensus mechanism.

Keywords

Semantic infrastructure, Symbiotic approach, DApp, SOCMINT, OSINT.

Two one-sided generalized inverses and their applications in bilinear models

Xiaomi Hu

Wichita State University, Kansas, USA

Abstract

Two one-sided generalized inverses of matrices are defined in this talk. The existence, uniqueness and the analytical forms through Moore-Penrose inverses are obtained. The two inverses play useful roles in establishing matrix projections onto linear spaces for matrices under a general inner product system. These results find applications in bilinear statistical models for random matrices in expressing general least square estimators for parameter matrices, in deriving sufficient and necessary conditions for the estimability of parameter matrices and in presenting best linear unbiased estimators (BLUEs) for parameter matrices.

Keywords

Generalized inverses, Matrix projections, Bilinear models for random matrices, Estimability for parameter matrices, Best linear unbiased estimators for parameter matrices.

References

[1] Hu, X. (2021). Matrix spaces and ordinary least square estimators in linear models for random matrices. *Submitted to Commun. Stat. Theory Methods.*

[2] Pan, J.X. and Fang, K.T. (2002). *Growth Curve Models and Statistical Diagonostics.* New York: Springer.

[3] Penrose, R. (1955). Generalized inverse for matrices. *Proc. Camb. Phil. Soc. 51*, 416–413.

[4] Penrose, R. (1956). On best approximate solutions of linear matrix equations. *Proc. Camb.Phil. Soc. 52*, 17–19.

[5] Pringle, R.M. and Rayner, A.A. (1971). *Generalized Inverse Matrices with Applications to Statistics.* London: Griffin.

[6] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications, 2nd Ed.* New York: Wiley.

Testing independence under a block compound symmetry covariance structure

Katarzyna Filipiak[1], Mateusz John[1] and Daniel Klein[2]

[1] Poznań University of Technology, Poland
[2] P. J. Šafárik University, Košice, Slovakia

Abstract

The goal of this article is to test the hypothesis related to the independence of features between any two repeated measures in a block compound symmetry structure under the doubly multivariate normal model.
The Rao score and Wald test statistics are determined and the characteristic function of the likelihood ratio test statistic is presented. For all of these test statistics, the asymptotic distributional properties are compared using simulation studies, and the robustness of the empirical distributions is considered. Furthermore, for power analysis purpose, the Kullback-Leibler divergence is proposed to measure discrepancy between hypotheses and the power of each mentioned tests, as well as F-test and Roy's largest root test, is studied. Finally, all mentioned tests are applied to a real data example

Keywords

Doubly multivariate model, Block compound symmetry, Rao score test, Wald test, Likelihood ratio test, Roy's largest root test, Independence, Entropy loss function, Power.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the contract no. APVV-17-0568 (Daniel Klein).

Comparison and assessment of shrinkage methods in case of multicollinearity problem

Şevval Kılıçoğlu and Fatma Yerlikaya-Özkurt

Atılım University, Ankara, Turkey

Abstract

Nowadays, the data analysis and interpretation are very important in many fields of science. One of the most preferred methods in data analysis is linear regression due to its simplicity to interpret and ease of application. Linear regression models are used to explain the relationship between a dependent variable and one or more independent variables. One of the assumptions accepted while obtaining linear regression models is that there is no correlation between the independent variables in the model which refers to absence of multicollinearity. As a result of multicollinearity, the variance of the parameter estimates will be high and this reduces the accuracy and reliability of the linear model.

The most powerful and preferred methods for eliminating the multicollinearity problem are shrinkage methods. Shrinkage methods aim to handle the multicollinearity problem by minimizing the variance of the estimators in the model. Ridge Regression, LASSO, and Elastic-Net are well-known and popular shrinkage methods. These methods converge the values of the coefficients of the variables in the model to zero or very close to zero values. In this study, mentioned shrinkage methods were applied to different simulated data sets with different characteristics and also three real-world data sets. After all models created, some performance criteria were calculated for each method to determine which method gives better results in the data set in which characters. Based on performance results, the methods were compared with TOPSIS, which is one of the Multi-Criteria Decision Making Methods, and the order of preference was determined for each data set.

Keywords

Multicollinearity, Shrinkage methods, Ridge regression, LASSO, Elastic Net, Multi criteria decision making, TOPSIS.

References

[1] Hoerl, A.E., and Kennard, R.W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics 12(1)*, 55–67.

[2] Jato-Espino, D., Castillo-Lopez, E., Rodriguez-Hernandez, J. and Canteras-Jordana, J.C. (2014). A review of application of multi-criteria decision making methods in construction. *Autom. Constr. 45*, 151–162.

[3] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B 58(1)*, 267–288.

[4] Uyanık, G.K. and Güler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Soc. Behav. Sci. 106*, 234–240.

[5] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol. 67(2)*, 301–320.

Association of suicide attempts with temperature by different humidity

Seunghyeon Kim and Eunsik Park

Chonnam National University, Gwangju, Korea

Abstract

The effect of temperature on suicide is clear. Many countries around the world have investigated the effect of temperature on suicide or suicide attempts and found that short-term exposure to high temperatures can be a trigger in common. However, the results by season differed across countries. The study in Israel showed that the effect of temperature and suicide attempts was significant only in summer, and the study in Korea was only significant in winter. We hypothesized that these seasonal differences are due to humidity. Therefore, this study investigated the modification of temperature effects on suicide attempts by humidity. We applied a time-stratified case-crossover design to control for seasonality and individual differences. We combined conditional Poisson regression with distributed lag non-linear models (DLNM) to explore a lag pattern and to obtain risk ratios (RR) and their 95% confidence intervals(CI) for the effects of the 80th percentile of temperature relative to the 10th percentile temperature. Then, interaction terms between relative humidity quartiles and temperature were generated to evaluate the modification of temperature effects on suicide attempts by relative humidity. Between 2014 and 2018, the count of suicide attempts is 39,803. The relative risk at the 80th percentile minimum temperature for suicide attempts was 1.09 (95% CI: 1.03, 1.15) compared with the risk at the 10th percentile minimum temperature at lag 1. In the case of maximum temperature, the relative risk was 1.12 (95% CI: 1.03, 1.22) at lag 0. The interaction effects between relative humidity and temperature effect were only found in the lowest humidity quantiles. This may explain why the effect of seasonal temperature varies from country to country. It also emphasizes that the association between temperature and suicide attempts is clear at low humidity. However, the study needs to be conducted in more countries to confirm the effect of temperature on humidity.

Keywords

Suicide attempts, Temperature, Relative humidity, Case cross-over design, Distributed lag non-linear model.

References

[1] Kalkstein, A.J., Belorid, M., Dixon, P.G., Kim, K.R. and Bremer, K.A. (2019). Seasonal variations in temperatureâĂŞsuicide associations across South Korea. *Weather, Climate, and Society 11(4)*, 731–739.

[2] Kubo, R., Ueda, K., Seposo, X., Honda, A. and Takano, H. (2021). Association between ambient temperature and intentional injuries: a case-crossover analysis using ambulance transport records in Japan. *Science of the total environment 774*, 145511.

[3] Gasparrini, A., Armstrong, B. and Kenward, M.G. (2010). Distributed lag nonlinear models. *Statistics in medicine 29(21)*, 2224–2234.

[4] Yarza, S., Vodonos, A., Hassan, L., Shalev, H., Novack, V. and Novack, L. (2020). Suicide behavior and meteorological characteristics in hot and arid climate. *Environmental research 184*, 109314.

Likelihood ratio test for testing covariance structures from commutative quadratic subspace

Katarzyna Filipiak[1], Yuli Liang[2] and Mateusz John[1]

[1] Poznań University of Technology, Poland
[2] Linnaeus University, Växjö, Sweden

Abstract

In this article, we address the problem of testing hypotheses about covariance structures from commutative quadratic subspace. The null distributions are established for the corresponding likelihood ratio test statistics. The application of the results is illustrated by both a simulation study and a real-life data example.

Keywords

Block compound symmetry, Beta random variables, Doubly multivariate model, Likelihood ratio test.

Skewness and kurtosis of mean-variance normal mixtures

Nicola Loperfido

Università degli Studi di Urbino "Carlo Bo", Urbino, Italy

Abstract

Mean-variance mixtures of normal distributions are very flexible distributions. They can model many nonnormal features, as for example skewness, kurtosis, or multimodality. Special cases include mixtures of two normal distributions with proportional covariance matrices and scale mixtures of normal distributions. This paper investigates the skewness and the kurtosis of multivariate mean-variance normal mixtures. Firstly, it derives the analytical forms of best-known measures of multivariate skewness and kurtosis, thus highlighting some of their shortcomings. Secondly, it applies these results to portfolio optimization, model-based clustering, invariant coordinate selection, normalizing linear transformations, projection pursuit and normality testing. The practical relevance of theoretical results in the paper is assessed using both real and simulated datasets.

Keywords

Kurtosis, Mixture model, Projection pursuit, Skewness.

References

[1] Lee, S.X. and McLachlan, G.J. (2019). *On mean and/or variance mixtures of normal distributions. In Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society*, 117–127. Cham: Springer.

[2] Malkovich, J.F. and Afifi, A.A. (1973). On Tests for Multivariate Normality. *J. Amer. Statist. Ass. 68*, 176–179.

[3] Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika 57*, 519–530.

[4] Mòri, T.F., Rohatgi, V.K. and Székely, G.J. (1993). On multivariate skewness and kurtosis. *Theory Probab. Appl. 38*, 547–551.

Extrapolation of random fields via level sets

Vitalii Makogin

Ulm University, Germany

Abstract

The literature on the inter- or extrapolation as well as prediction of random processes and fields is huge. In the infinite variance case, however, the approaches are tailored to specific classes of processes or fields under consideration and the general framework for the extrapolation of heavy-tailed random fields is still missing. We try to fill this gap by noting that two random fields are, in a sense, similar if their level (or excursion) sets are similar. To be more precise, two random fields modeling some feature with the same structure of excursions have the same total amount of this feature exceeding each level over a fixed time interval or a spatial domain. This is certainly of interest for practical applications to insurance (with the feature being the claim size), environmetrics (e.g. for the amount of environmental pollution or radiation), etc. In our approach, similarity is measured by the expected volume of the symmetric difference of the level sets. It is sometimes also called expected distance in measure.
We use this concept of excursion sets for the extrapolation of stationary random fields. Doing so, we define excursion sets for the field and its linear predictor, and then minimize the expected volume of the symmetric difference of these sets under the condition that the univariate distributions of the predictor and of the field itself coincide. We illustrate the new approach on Gaussian random fields.

Keywords

Stationary random field, Gaussian random field, Extrapolation, Linear prediction, Excursion, Level set.

A generalized linear failure rate distribution as a parametric model for some medical studies

Mahmoud Mansour

Al-Azhar University, Nasr City, Cairo, Egypt

Abstract

In medical studies, some patients exposed to a treatment program may leave that program for unknown reasons. This experimental situation results in the loss of some data and in turn affect the decision to be made based on the study aimed at evaluating the quality of the treatment program. The issue is how to compensate for this loss of data and hence make the right decision. A multiply-hybrid censored scheme is applied through a parametric model called Generalized Linear Failure Rate (GLFR) to predict the missed data. Some statistical Bayesian approaches will be applied to the survival times for a group of patients diagnosed with a kind of leukemia.

Keywords

Statistical inference, Generalized linear failure rate, Bayesian approach, Censored data, Two sample prediction, Assessment of treatment methods.

References

[1] Danish, M.Y., Arshad, I.A. and Aslam, M. (2018). Bayesian inference for the randomly censored Burr-type XII distribution. *Journal of Applied Statistics 45(2)*, 270–283.

[2] Jeon, Y.E. and Kang, S.B. (2020). Estimation for the half-logistic distribution based on multiply Type-II hybrid censoring. *Physica A 550*. Available online https://doi.org/10.1016/j.physa.2020.124501.

[3] Mahmoud, M.A.W., Ramadan, D.A. and Mansour, M.M.M. (2020). Estimation of lifetime parameters of the modified extended exponential distribution with application to a mechanical model. *Communications in Statistics - Simulation and Computation*. Available online https://doi.org/10.1080/03610918.2020.1821887.

[4] Park, S. and Balakrishnan, N. (2012). A very flexible hybrid censoring scheme and its Fisher information. *Journal of Statistical Computation and Simulation 82*, 41–50.

Improved estimators of linearly structured covariance matrices

Augustyn Markiewicz and Adam Mieldzioc

Poznań University of Life Sciences, Poland

Abstract

In many applications, the positive definite and well-conditioned estimator of the covariance matrix or its inverse is required. The standard method of determining this estimator is maximum likelihood estimation. Its estimation becomes complicated when it has a linear structure. The maximum likelihood estimator, in this case, can not have an explicit form and its derivation is challenging and time-consuming. [1] proposed an explicit estimation procedure based on the orthogonal projection of the sample covariance matrix on a given linear structure. The resulting estimator has good statistical properties, but this method does not ensure its positive definiteness and well-conditioning. An improvement of this method is proposed that leads to the fast estimation of linearly structured covariance matrix by adapting the shrinkage method. The new estimator and maximum likelihood estimator are compared with respect to several risk functions. Their statistical properties are similar, but the new method is faster and therefore it can be recommended.

Keywords

Linearly structured covariance matrix, Orthogonal projection, Shrinkage estimation, Maximum likelihood estimation.

References

[1] Ohlson, M. and von Rosen, D. (2010). Explicit estimators of parameters in the Growth Curve model with linearly structured covariance matrices. *Journal of Multivariate Analysis 101*, 1284–1295.

Matrix variate generalized laplace distributions

Tomasz J. Kozubowski[1], Stepan Mazur[2,3] and Krzysztof Podgórski[4]

[1] University of Nevada, USA
[2] Örebro University, Sweden
[3] Linnaeus University, Växjö, Sweden
[4] Lund University, Sweden

Abstract

The generalized asymmetric Laplace (GAL) distribution, also known as the *variance/mean-gamma* model, is a popular flexible class of distributions that can account for peakedness, skewness, and heavier than normal tails, often observed in financial or other empirical data. We consider extensions of the GAL distribution to the matrix variate case, which arise as covariance mixtures of matrix variate normal distributions. Two different mixing mechanisms connected with the nature of the random scaling matrix are considered, leading to what we term matrix variate GAL distributions of Type I and II. While Type I matrix variate GAL distribution has been studied before, there is no comprehensive account of Type II in the literature, except for their rather brief treatment as a special case of matrix variate generalized hyperbolic distributions. With this work we fill this gap, and present an account for basic distributional properties of Type II matrix variate GAL distributions. In particular, we derive their probability density function and the characteristic function, as well as provide stochastic representations related to matrix variate gamma distribution. We also show that this distribution is closed under linear transformations, and study the relevant marginal distributions. In addition, we also briefly account for Type I and discuss the interconnections with Type II. We hope that this work will be useful in the areas where matrix variate distributions provide an appropriate probabilistic tool for three-way or, more generally, panel data sets, which can arise across different applications.

Keywords

Covariance mixture of Gaussian distributions, Distribution theory, Generalized Laplace distribution, MatG distribution, Matrix variate distribution, Matrix variate gamma distribution, Matrix gamma-normal distribution, Matrix variate t distribution, Normal variance-mean mixture, Variance gamma distribution.

A new proposal for robust estimation of the extremal index

M. Cristina Miranda[1], Manuela Souto de Miranda[1] and M. Ivette Gomes[2]

[1] Universidade de Aveiro, Portugal
[2] Universidade de Lisboa, Portugal

Abstract

The extremal index is a parameter defined in the framework of Extreme Value Theory', which measures the degree of dependence among exceedances above high fixed thresholds. When the extremal index exists and those exceedances occur in clusters, the extremal index (*EI*) is related to the dimension of the clusters and, in the limit distribution, it coincides with the reciprocal of the mean clusters dimension. There are different *EI* estimators according to the method used in the identification of those clusters. Some of them use the sample mean in the estimation of the cluster dimension. The present proposal is based on the "runs" estimator. It considers a negative binomial as the limit distribution of the number of exceedances that occur before a non-exceedance observation. Then, the *EI* is estimated by the reciprocal of the constant term of a negative binomial regression. The procedure makes use of robust estimators for counting processes, with known properties. Thus the negative binomial distribution is integrated in the estimation of the mean, while small deviations from the assumptions are controlled, including the occurrence of atypical cluster size values. A simulation study explores and compares the present robust proposal with other estimators.

Keywords

Extremal index, Robustness, Estimation, Negative binomial regression, Simulation.

References

[1] Aeberhard, W.H., Cantoni, E. and Heritier, S. (2014). Robust inference in the negative binomial regression model with an application to falls data. *Biometrics 70(4)*, 920–31.
[2] Gomes, M.I., Miranda, M.C. and Souto de Miranda, M. (2020). A Note on Robust Estimation of the Extremal Index in Springer *Proc Math Stat. 339*, 213–25.
[3] Maronna, R.A., Martin, R.D., Yohai, V.J. and Salibián-Barrera, M. (2019). *Robust statistics : theory and methods (with R)*. Wiley.
[4] Weissman, I. and Novak, S.Y. (1998). On blocks and runs estimators of the extremal index. *J Stat Plan Inference. 66(2)*, 281–288.

Discrepancy measures between structured covariance matrices

Katarzyna Filipiak[1], Daniel Klein[2] and Monika Mokrzycka[3]

[1] Poznań University of Technology, Poland
[2] P. J. Šafárik University, Košice, Slovakia
[3] Polish Academy of Sciences, Poznań, Poland

Abstract

In the literature various tests for covariance structures have been proposed, however, the method of measuring discrepancy between two distributions which differ in covariance matrix structures related to the study of power of the test remains an open problem. The aim of this paper is to verify the properties of the power of the test due to various discrepancy measures: simple one parameter discrepancies or more complex discrepancies based on minimization of Frobenius norm or entropy/quadratic loss functions. The criterion of the choice of the most suitable measure of discrepancy is based on increasing property of power function with respect to discrepancy.
In the paper the power of likelihood ratio and Rao score tests are considered, and the basic hypothesis is related to separable structure of the observation matrix under doubly multivariate model, however, presented results can be also applied to more general or more detailed covariance structures.

Keywords

Covariance structures, Power of the test, Frobenius norm, Entropy loss function, Quadratic loss function, Likelihood ratio test, Rao score test.

Tetra-alphabetic hypercubes and atmosphere sampling

Paulo Monteiro[1], João Tiago Mexia[2], M. Manuela Oliveira[1], Francisco Carvalho[2,3] and Cristina Andrade[3,4]

[1] Universidade de Évora, Portugal
[2] Universidade Nova de Lisboa, Portugal
[3] Instituto Politécnico de Tomar, Portugal
[4] Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal

Abstract

Tetra-alphabetic hypercubes use four distinct alphabets with orthogonality between them. These hypercubes enable balanced four-dimensional sampling. In the case of atmospheric studies, we will consider the four dimensions as latitude, longitude, hight and time.
An application will be presented, showing the advantages of our approach to sampling obtained from atmospheric data and it's consequences for the study under huge amounts of data, computational procedure and quickness in obtaining results.

Keywords

Atmosphere, Balanced sampling, Tetra-alphabetic hypercubes.

References

[1] Cochran, W.G. (1991). *Sampling Techniques, 3rd Edition*. Wiley.
[2] Dey, A. and Mukerjee, R. (1999). *Fractional Factorial Plans*. Wiley-Interscience.
[3] Hinkelmann, K. and Kempthorne, O. (2005). *Design and Analysis of Experiments: Advanced Experimental Design, vol. 2*. Wiley.
[4] Mukerjee, R. and Wu, C.F.J. (2006). *A Modern Theory of Factorial Design*. Springer Series in Statistics.

Analysis of changes in annual precipitation patterns in Alentejo region using log-linear models

Vítor Augusto and Elsa Moreira

Nova School of Sciences and Technology, Caparica, Portugal

Abstract

Climatic changes are a topic of extreme importance nowadays. With the aim of adapting agriculture in Alentejo to climate changes, this work intended to find statistically significant differences in the intra-annual and inter-annual cycles of precipitation in Alentejo, Portugal, over the last 40 years. To do so, precipitation data from each location in Alentejo were divided into four decades and grouped into contingency tables in order to fit log-linear models with two categories: year and month. ANOVA-type tables were obtained with residual deviations for the factors month, year and interaction, which allowed us to know their statistical significance in the model. Then, backward elimination method was applied to try to reduce the parameters of the models referring to the months that were less relevant to explain the variability of precipitation. In the end, we were able to conclude that in the oldest decade there was more intra-annual variability of precipitation, which could be interpreted as trend towards smoothing out the differences in precipitation between the months of the year. Furthermore, with regard to the inter-annual variability, a cyclical behaviour emerged when comparing the 4 decades.

Keywords

Climatic changes, Contingency tables, Backward elimination, Residual deviance.

References

[1] Agresti, A. (1990). *An Introduction to Categorical Data Analysis*. New York: John Willey & Sons.

[2] Dunn, P.K. and Smyth, G.K. (2018). *Generalized Linear Model with Examples in R*. New York: Springer.

[3] Everitt, B. (1977). *The analysis of Contingency Tables*. London: Chapman & Hall.

[4] Moreira, E., Mexia, J.T., Pereira, L. (2012). Are drought occurrence and severity aggravating? A study on SPI drought class transitions using log-linear models and ANOVA-like inference. *Hydrology and Earth System Sciences 16*, 3011–3028.

[5] Moreira, E., Mexia, J.T., Pereira, L. (2013). Assessing homogeneous regions relative to drought class transitions using an ANOVA-like inference. Application to Alentejo, Portugal. *Stochastic Environmental Research and Risk Assessment 27(1)*, 183–193.

A note on the restricted enet estimators

Mina Norouzirad[1], Filipe. J. Marques[1], Danial Mazarei[2]

[1] Nova School of Science and Technology, Caparica, Portugal
[2] Persian Gulf University, Bushehr, Iran

Abstract

In a regression model, information is always critical to estimate accurately. The information is generally limited to sample data, and no prior knowledge of the model's parameters is assumed. Sparsity and multicollinearity are examples of theoretical knowledge that have been utilized to develop penalized estimators. However, other sources of information, such as expert suggestions, previous experiments, and studies, can also contribute non-sample information. This study aims to combine all available information to construct a new estimator called the restricted elastic net estimator. Incorrect information affects the accuracy of a prediction, but accurate information reduces prediction error. This claim was supported by using correct information to develop a more precise estimate and collect evidence to prove it. Simulated research and real-data applications demonstrated this.

Keywords

Double shrinking, Enet estimator, Prior information, Restricted Enet estimator.

References

[1] Saleh, A.K.M.E., Arashi, M., Saleh, R.A. and Norouzirad, M. (2022). *Rank-Based Methods for Shrinkage and Selection: With Application to Machine Learning*. John Wiley & Sons.
[2] Norouzirad, M. and Arashi, M. (2018). Preliminary test and Stein-type shrinkage LASSO-based estimators. *SORT-Statistics and Operations Research Transactions 1(1)*, 45–58.
[3] Saleh, A.K.M.E. (2006). *Theory of preliminary test and stein-type estimation with applications*. John Wiley & Sons.
[4] Zou, H. and Hastie T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 67(2)*, 301–320.

Approximate reference priors for Gaussian random fields

Victor De Oliveira[1] and Zifei Han[2]

[1] The University of Texas at San Antonio, Texas, USA
[2] University of International Business and Economics, Beijing, China

Abstract

When modeling spatially correlated data using Gaussian random fields, exact reference priors for the model parameters have been recommended for objective Bayesian analysis. But their use in practice is hindered by its complex formulation and the associated computational costs. In this work, we propose a new class of default prior distributions for the parameters of Gaussian random fields that approximate exact reference priors. It is based on the spectral representation of stationary random fields and their spectral density functions. These approximate reference priors maintain the major theoretical advantages of exact reference priors, but at a much lower computational cost. Unlike the situation for exact reference priors, we show that the marginal prior of the range parameter in the Matern correlation family is always proper, regardless of the mean function or degree of smoothness of the correlation function, and also establish the propriety of the joint reference posterior of the model parameters. Finally, an illustration is provided with a spatial data set of lead pollution in Galicia, Spain.

Keywords

Bayesian analysis, Default prior, Geostatistics, Spectral representation.

References

[1] Berger, J., De Oliveira, V. and Sansó, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association 96*, 1361–1374.
[2] De Oliveira, V. (2007). Bayesian analysis of spatial data with measurement error. *The Canadian Journal of Statistics 35*, 283–301.
[3] Handcock, M. and Stein, M. (1993). A Bayesian analysis of kriging. *Technometrics 35*, 403–410.
[4] Muré, J. (2021). Propriety of the reference posterior distribution in Gaussian process modeling. *The Annals of Statistics 49*, 2356–2377.
[5] Paciorek, C. (2007). Bayesian smoothing with Gaussian processes using Fourier basis functions in the spectralGP package. *Journal of Statistical Software 19*, 1–38.
[6] Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag.
[7] Yaglom, A. (1987). *Correlation Theory of Stationary and Related Random Functions I: Basic Results*. Springer-Verlag.

Fuzzy clustering approach to quantification of allostatic load

J.A. Pereira[1,2,3], F. Abreu[1], L. Mendes[1] and T. Oliveira[2,3]

[1] Universidade do Porto, Portugal
[2] CEAUL, Lisboa, Portugal
[3] Universidade Aberta, DCeT, Portugal

Abstract

Purpose: Allostatic load (AL) refers to the cumulative burden of chronic stress and life events and results in multi-systemic physiological dysregulation. Periodontal Disease (PD) has been associated with various systemic diseases and is impacted by metabolic dysregulation. The aim of this research is to assess the relationship between AL and PD using unsupervised machine learning fuzzy methods.
Methods and Results: Data from the National Health and Nutrition Examination Survey (NHANES) 2011 was used. AL was measured using eleven biomarkers representing cardiovascular, inflammatory, and metabolic system functioning outcomes. A total of 1414 US adults aged 35 years and older were allocated to two fuzzy clusters, using the Gustafson, Kessel, and Babuska c-means type algorithm (GKB-FkM), the fuzzy k-means clustering algorithm. The cluster 1 presented more advantageous values for the allostatic load surrogate biomarkers. In both clusters, the membership degrees (MD) varied from 0.5 and 1.0, with an average of 0.7. The PD parameters' were compared between both clusters using GAMLSS models, yielding statistically significant differences ($p < 0.05$) for pocket probing depth (PPD) mean and maximum and clinical attachment loss (CAL) rate. The correlation coefficients between PD parameters and cluster 1 MD's ranged from -0.06 and -0.11, being statistically significant. The association of PPD mean with cluster 1 MD's statistical significance did hold up after adjustment for age and gender.
Conclusion: The latent nature of AL together with the absence of an universally accepted AL score poses major difficulties when classification of individuals is needed and to correlate their allostatic burden with other conditions. We propose to tackle this issue using fuzzy clustering methods in combinations with GAMLSS models. This approach allowed us to find an association between AL and PD by measuring individuals AL through the membership grade to a cluster that's hold after adjustment for age and gender. This methodology appears to be promising to deal with variables that results from a complex combination of surrogate endpoints, and which aggregation is difficult or impossible.

Keywords:

Allostatic load, Pperiodontal health, Fuzzy clustering, GAMLSS.

References

[1] Guidi, J., Lucente, M., Sonino, N. and Fava, G.A.(2021). Allostatic Load and Its Impact on Health: A Systematic Review. *Psychother. Psychosom 90*, 11–27.

[2] Kim, J., Amar, S. (2006). Periodontal disease and systemic conditions: a bidirectional relationship. *Odontology 94(1)*, 10–21.
[3] Chaudhuri, A. (2019). *Fuzzy Machine Learning: Advanced Approaches to Solve Optimization Problems*. De Gruyter.

Permissible covariance structures for simultaneous retention of BLUEs

Simo Puntanen

Tampere University, Finland

Abstract

In this talk we consider the partitioned linear model

$$\mathscr{M}_{12}(\mathbf{V}_0) = \{\mathbf{y},\, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2,\, \mathbf{V}_0\}$$

and the corresponding small model

$$\mathscr{M}_1(\mathbf{V}_0) = \{\mathbf{y},\, \mathbf{X}_1\boldsymbol{\beta}_1,\, \mathbf{V}_0\}.$$

Following [1, Sec. 5.2] we can characterize the set $\mathscr{V}_{12}$ of nonnegative definite matrices $\mathbf{V}$ such that every representation of the best linear unbiased estimator, BLUE, of $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ under $\mathscr{M}_{12}(\mathbf{V}_0)$ remains BLUE under $\mathscr{M}_{12}(\mathbf{V})$. Correspondingly, we can characterize the set $\mathscr{V}_1$ of matrices $\mathbf{V}$ such that every BLUE of $\boldsymbol{\mu}_1 = \mathbf{X}_1\boldsymbol{\beta}_1$ under $\mathscr{M}_1(\mathbf{V}_0)$ remains BLUE under $\mathscr{M}_1(\mathbf{V})$. In this talk we focus on the mutual relations between the sets $\mathscr{V}_1$ and $\mathscr{V}_{12}$.
This talk is based on co-operation with Stephen J. Haslett, Jarkko Isotalo and Augustyn Markiewicz.

Keywords

Best linear unbiased estimator, BLUE, Covariance matrix, Equality of the BLUEs, OLSE, Partitioned linear model.

References

[1] Rao, C.R. (1971). Unified theory of linear estimation. *Sankhyā Ser. A, 33*, 371–394. [Corrigenda (1972): 34, p. 194 and p. 477]

Time series forecasting using ensemble and hybrid methodologies

Paulo Canas Rodrigues

Federal University of Bahia, Salvador, Brazil

Abstract

Time series forecasting plays a key role in areas such as energy, environment, economy, and finances. Hybrid methodologies, combining the results of statistical and machine learning methods, have become popular for time series analysis and forecasting, as they allow researchers to compensate the limitations of one approach with the strengths of the other, and combine them into new frameworks while improving forecasting accuracy. In this class of methods, algorithms for time series forecasting are applied sequentially, i.e., the second model can be applied to the residuals that were not captured by the first, by considering that the observed data is a combination of linear and nonlinear components. Another category of methods for time series forecasting is the ensemble methods that are the result of the weighted average of individual forecasts from several methods. In this talk, I will discuss several strategies for time series forecasting, including ensemble and hybrid methodologies, with application to electricity load forecasting and to PM10 (inhalable particles, with diameters that are generally 10 micrometers and smaller) forecasting.

Keywords

Time series forecasting, Singular spectrum analysis, Neural networks, Ensemble methods, Hybrid methodologies.

References

[1] Cordova, C.H., Portocarrero, M.N.L., Salas, R., Rodrigues, P.C. and López- Gonzales, J.L. (2021). Air Quality Assessment and Pollution Forecasting using Recurrent Artificial Neural Networks in Metropolitan Lima-Peru. *Scientific Reports 11*, 24232.

[2] Encalada-Malca, A.A., Cochachi-Bustamante, J.D., Rodrigues, P.C., Salas, R. and López-Gonzales, J.L. (2021). A spatio-temporal visualization approach to the exploration of PM10 concentration data in Metropolitan Lima. *Atmosphere 12*, 609.

[3] Sulandari, W., Subanar, Lee, M.H. and Rodrigues, P.C. (2020). Indonesian electricity load forecasting using singular spectrum analysis. *Energy 190*, 116408.

[4] Sulandari, W., Subanar, Lee, M.H. and Rodrigues, P.C. (2020). Time series forecasting using singular spectrum analysis, fuzzy systems and neural networks. *MethodsX 7*, 101015.

FGLS estimator: a robust approach with panel data

Anabela Rocha and M. Cristina Miranda

Universidade de Aveiro, Portugal

Abstract

Econometric and financial studies are frequently related to panel data. These type of data arise whenever it is relevant to follow up the behaviour of some variables among a set of companies or financial institutions, for example. Panel data models apply to repeated observations of the same set of units over different time moments. They allow to identify the existence of non-observable effects that would be unnoticed with other models, like an individual or firm effect characteristic. If a set of financial indicators is recorded for the same set of chosen firms for different moments in time, it might be interesting to know if the results are affected by some firm feature, constant in time. Under suitable conditions, the Feasible Generalized Least Squares (FGLS) procedure is applied to obtain the parameter estimators of panel data models. These estimators may be seriously affected by the existence of outliers, which frequently appear in econometric data. Robust estimation presents a solution as it is less affected by the occurrence of such atypical observations but it is not yet commonly seen in economical and financial empirical studies. In this paper, the authors aim to contribute to the development of robust estimation methodologies with panel data. They present a robust version of the traditional FGLS estimator, the Robust FGLS (RFGLS), and compare the performance of the two estimators using a wellknown set of economical data. The study is complemented with a Monte Carlo simulation assessing the estimator properties under the presence of some undesired scenarios, so common in economical and financial areas.

Keywords

FGLS, Panel data, Robust estimation, Simulation.

References

[1] Baltagi, B.H. (2001). *Econometric Analysis of Panel Data*. John Wiley & Sons.

[2] Bramati, M.C. (2007). Robust Estimators for the Fixed Effects Panel Data Model. *Econometrics Journal 10(3)*, 521–540.

[3] Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

[4] Maronna, R.A., Martin, R.D., Yohai, V.J. and Salibián-Barrera, M. (2019). *Robust statistics: theory and methods (with R)*. Wiley.

Principal component analysis and singular multivariate regression

Dietrich von Rosen[1,2]

[1] Swedish University of Agricultural Sciences, Uppsala, Sweden
[2] Linköping University, Sweden

Abstract

Data reduction is important. Multivariate linear models (MANOVA) and bilinear models (Growth Curve model) are treated when the dispersion matrix is unknown but assumed to be singular. This can be used as a tool for reducing the size of the observed vectors. The approach leads to a basic model for principal component analysis. Estimators will be presented and some comments on the importance of working with observed data belonging to appropriate subspaces is discussed. If data do not belong to these subspaces data have to be preprocessed by applying certain projections.

Keywords

Multivariate linear models, Principal component analysis, Singular dispersion matrix.

Joining iso-structured models with commutative orthogonal block structure

Carla Santos[1,2], Cristina Dias[2,3], Célia Nunes[4] and João T. Mexia[2]

[1] Polytechnic Institute of Beja, Portugal
[2] Nova University of Lisbon, Portugal
[3] Polytechnic Institute of Portalegre, Portugal
[4] University of Beira Interior, Covilhã, Portugal

Abstract

A model with commutative orthogonal block structure (COBS) is a linear mixed model whose variance-covariance matrix is a linear combination of known pairwise orthogonal projection matrices that add up to the identity matrix, and commutes with the orthogonal projection matrix on the space spanned by the mean vector. COBS, as particular class of the models with orthogonal block structure, arose in order to obtain optimal estimation for variance components of blocks and contrasts of treatments. Resorting to their algebraic structure, we study COBS and the operation of models joining. Since joining COBS originates a new COBS, ensuring that the conditions for the good properties of the estimators are preserved, we explore performing the operation of models joining with iso-stuctured COBS, that is, with COBS with identical space spanned by their mean vectors and having covariance matrices that are linear combinations of the same pairwise orthogonal projection matrices.

Keywords

Best linear unbiased estimator, Mixed model, Jordan algebra, Operation with models.

Acknowledgments
This work was partially supported by national founds of FCT- Fundação para a Ciência e Tecnologia (Foundation for Science and Technology), Portugal, under UIDB/00297/2020 and UIDB/00212/2020.

References

[1] Jordan, P., von Neumann, J. and Wigner, E. (1934). On an algebraic generalization of the quantum mechanical formulation. *Annals of Math. 35(1)*, 29–64.
[2] Nelder, J.A. (1965a) The analysis of randomized experiments with orthogonal block structure I. Block structure and the null analysis of variance. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 283*, 147–162.
[3] Nelder, J.A. (1965b) The analysis of randomized experiments with orthogonal block structure II. Treatment structure and the general analysis of variance. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 283*, 163–178.
[4] Santos, C., Nunes, C., Dias, C. and Mexia, J.T. (2017). Joining models with commutative orthogonal block structure. *Linear Algebra and its Applications 517*, 235–245.

[5] Zmyślony, R. (1978). A characterization of best linear unbiased estimators in the general linear model. *Lecture Notes in Statistics 2*, 365–373.

A new approach to measure divergence using STATIS. An application to Malapportionment

Antonio Seijas-Macías[1,3], Amílcar Oliveira[2,3] and Teresa A. Oliveira[2,3]

[1] Universidade da Coruña, Spain
[2] Universidade Aberta, Portugal
[3] Universidade de Lisboa, Portugal

Abstract

This paper analyses the measure of Malapportionment. The most common method to analyse the discrepancy between the share of legislative seats and the share of the population for the constituencies is the Loosemore-Hanby index-based. But recently some other proposals have been presented: the max-min ratio, the Gini index and the α-divergence.
The STATIS method is a data analysis technique that handles three-way class data as a set of matrices. A generalisation for the four-way multiblock method (STATIS-4) exists for studying the links between more than two sets of data tables.
In this paper, we consider the application of STATIS to measure malapportionment into a country or parliament. We compare the results with some other traditional measures to establish the benefits of this approach.

Keywords

Exploratory analysis, Disproportionality measure, STATIS, Gini Index.

References

[1] Kamahara, Y. and Kasuya, Y. (2014). The State of Malapportionment in theWorld: One Person, One Vote?. *Research note prepared for delivery at the Annual Conference of the Australian Political Studies Association*, 1–31.
[2] Lavit, C., Escoufier, Y., Sabatier, R. and Raissac, P. (1994). The ACT (STATIS method). *Computational Statistics and Data Analysis 18*, 97–119.
[3] Sabatier, R. and Vivien, M. (2008). A new linear method for analysing four-way multi-blocks tables: STATIS-4. *Journal of Chemometrics 22*, 309–407.
[4] Stanimirova, I., Walczak, B., Massart, D.L., Simeonov, V., Saby, C.A. and Di Crescenzo, E. (2004). STATIS is a three-way method for data analysis. Application to environmental data. *Chemometrics and intelligent laboratory systems 73*, 219–233.
[5] Wada, J. and Kamahara, Y. (2018). Studying malapportionment using α-divergence. *Mathematics Social Science 93*, 77–89.

Scaling priors in two dimensions for intrinsic Gaussian Markov Random Fields

Maria-Zafeiria Spyropoulou and James Bentham

University of Kent, UK

Abstract

Intrinsic Gaussian Markov Random Fields (IGMRFs) can be used to induce conditional dependence in Bayesian hierarchical models. IGMRFs have both a precision matrix, which defines the neighbourhood structure of the model, and a precision, or scaling, parameter. Previous studies have shown the importance of selecting the prior of this scaling parameter appropriately for different types of IGMRF, as it can have a substantial impact on posterior results. Here, we focus on the two-dimensional case, where tuning of the parameter's prior is achieved by mapping it to the marginal standard deviation of a two-dimensional IGMRF. We compare the effects of scaling various classes of IGMRF, including an application to blood pressure data using MCMC methods.

Keywords

Priors, Intrinsic Gaussian Markov Random Fields, MCMC, Precision, Scaling, Two-dimensional problems.

References

[1] Sørbye, S.H. and Rue, H. (2014). Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics 8*, 39–51.

A robust scalar-on-function partial quantile regression

Müjgan Tez[1], Ufuk Beyaztas[1] and Han Lin Shang[2]

[1] Marmara University, Istanbul, Turkey
[2] Macquarie University, Sydney, Australia

Abstract

Scalar-on-function quantile regression is a powerful regression model to characterize the entire conditional distribution of a scalar response variable for a given functional predictor. Compared with the conditional mean regression-based scalar-on-function regression model, the scalar-on-function quantile regression is robust to outliers in the response variable. However, it is susceptible to outliers in the functional predictor (called leverage points). This is because the influence function of the regression quantiles is bounded in the response variable but unbounded in the predictor space. The leverage points may alter the eigenstructure of the predictor matrix, leading to poor estimation and prediction results. This study proposes a robust scalar-onfunction quantile regression method to robustly estimate the model parameters and produce reliable predictions in the presence of both outliers and leverage points. The proposed method is based on a functional partial quantile regression procedure. We propose a robust partial quantile covariance to obtain functional partial quantile components of the scalar-on-function regression model. In the functional partial quantile decomposition, the robustness is obtained by iteratively reweighting the functional predictor. After such a decomposition, the infinite-dimensional scalar-on-function quantile regression model is approximated in the finite-dimensional space by a quantile regression of the scalar response on the partial quantile scores. The estimation and prediction performance of the proposed method is evaluated by a series of Monte-Carlo experiments and an empirical data example, and the results are compared favorably with several existing methods.

Keywords

Functional data, Iteratively reweighting, Partial quantile covariance, Robust estimation.

Nonparametric approach for approximating the ruin probability of Sparre Andersen risk models

Atik Touazi[1,2] and Djamil Aissani[2]

[1] University of Algiers, Algeria
[2] University of Bejaia, Algeria

Abstract

Ruin probability is one of the main features of a risk model which is the probability that the insurer capital falls bellow the level zero. However, it cannot be found explicitly for many risk models except in rare cases. This is so-called Cramér-Lundberg risk model where claims follow an exponential distribution (see [2]). In such situation the stability question has been investigated for stochastic risk models in one-dimensional (Kalashnikov 2000; Enikeeva 2001) and in two-dimensional [3]. The strong stability method was developed in the beginning of the 1980s [1]. It is applicable to all models which can be represented by a Markov chain. The advantage of this method is its ability to realize a qualitative and quantitative analysis of some complex models. In contrast to other methods, we suppose that the perturbation of the transition kernel is small with respect to a certain norm in the operator space. Such a strict condition allows us to obtain better estimate of the ruin probability of the perturbed risk model (see [4, 5]). The aim of this work is to prove the applicability of the strong stability method of the Sparre Andersen risk model when the claims inter-occurrence times distribution function is general and unknown. In this case, we should use kernel density estimation methods to estimate the unknown density function of claims inter-occurrence times. Using simulation approach, we evaluate numerically the approximation error between the ruin probability in the Sparre Andersen model and the ruin probability in the Cramér-Lundberg model.

Keywords

Nonparametric statistics, Strong stability, Stochastic risk models, Simulation.

References

[1] Aissani, D. and Kartashov, N.V.(1983). Ergodicity and stability of Markov chains with respect to operator topology in the space of transition kernels. *Dokl. Akad. Nauk Ukr. Ser. A 11*, 3–5.
[2] Asmussen, S. and Albrecher, H. (2010). *Ruin probabilities.*Word Scientific.
[3] Benouaret, Z. and Aissani, D. (2010). Strong stability in a two-dimensional classical risk model with independent claims. *Scandinavian Actuarial Journal 2*, 83–93.
[4] Harfouche, Z. and Bareche, A. (2021). Semi-parametric approach for approximating the ruin probability of classical risk models with large claims. *Communication in Statistics - Simulation and Computation*. doi:10.1080/03610918.2021.1992636.
[5] Touazi, A., Benouaret, Z., Aissani, D. and Adjabi, S. (2017). Nonparametric estimation of the claim amount in the strong stability analysis of the classical risk model. *Insurance : Mathematics and Economics 74*, 78–83.

Estimating causal effects from complex longitudinal data via point effects of treatments

Xiaoqin Wang[1] and Li Yin[2]

[1] University of Gävle, Sweden
[2] Karolinska Institutet, Stockholm, Sweden

Abstract

Background and purpose: In many practices, one assigns a sequence of treatments to influence a certain outcome of interest. Between treatments, there exist the time-dependent covariates, which may have influences from earlier treatments and on subsequent treatments. From the complex longitudinal data, one aims to estimate the causal effect of a sequence of treatments on the outcome.
Challenges: Usually, one estimates the causal effect by modeling the standard parameters, i.e., the conditional mean of the outcome given all treatments and covariates. However, these parameters are essentially all different (null paradox). Furthermore, the dimension of the parameters is huge (curse of dimensionality).
Solutions: Instead of standard parameters, Wang and Yin estimated the causal effect by modeling the point effects of treatments in the sequence. The point effect is simply the effect of single-point treatment in single-point causal inference and its estimation is well studied.
Achievements: We have applied this method to a longitudinal study of COVID-19 progression during the first-wave pandemic and found that Swedish measure during the initial period have a long-term causal effect on the COVID-19 mortality as compared to the measures adopted by the other Nordic countries.

Keywords

Long-term causal effect, Short-term causal effect, Point effect, COVID-19 progression.

References

[1] Wang, X. and Yin, L. (2020). New G-Formula for the Sequential Causal Effect and Blip Effect of Treatment in Sequential Causal Inference. *Annals of Statistics 48*, 138–160.

Robust cyrptocurrency portfolio optimization by using MNN and NNNPO

Ilgim Yaman

Giresun University, Turkey

Abstract

The more important the gain is for the investors, the more important the risk is. Therefore, the portfolio optimization is a major problem in the finance world. Harry Markowitz had proposed standard portfolio optimization method in 1952. In terms of dealing with the fitness function of the portfolio optimization model, solving this quadratic optimization problem is not that simple. In 2019, we proposed nonlinear neural network for portfolio optimization (NNNPO) which solves that complex problem for the standard portfolio optimization problem. Nowadays crypto assets more common than stock exchange market, but cryptocurrencies do not meet Markowitz's assumption which is assets have to be normally distributed. In this study, cryptocurrency data taken in to account between July 5, 2018 to July 4, 2019. Firstly, to evaluate a portfolio return for risk, information ratio, sharpe ratio, sortino ratio calculated. Secondly, to avoid normality assumption, robust parameters calculated. Thirdly, the forecasts obtained from multiplicative neural networks (MNN). Then, forecasted data taken into nonlinear neural network algorithm for portfolio optimization problem to determine the proportion of the currencies in the selected portfolio. Finally, the return of selected portfolio compared with the real cryptocurrency data.

Keywords

Nonlinear neural network, Robust cyrptocurrency portfolio optimization, Multiplicative neural networks.

References

[1] Markowitz, H. (1952). The utility of wealth. *Journal of political Economy 60(2)*, 151–158.

[2] Egrioglu, E., Yolcu, U., Aladag, C.H. and Bas, E. (2015). Recurrent multiplicative neuron model artificial neural network for non-linear time series forecasting. *Neural Processing Letters 41(2)*, 249–258.

[3] Yaman, I., Dalkiliç , T.E. (2019). Portfolio selection based on a nonlinear neural network: An application on the Istanbul Stock Exchange (ISE30). *Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics 68(2)*, 1709–1723.

Spline based sparseness and smoothness for partially nonlinear model via C-fused lasso

Fatma Yerlikaya-Özkurt[1], Pakize Taylan[2] and Müjgan Tez[3]

[1] Atılım University, Ankara, Turkey
[2] Dicle University, Diyarbakır, Turkey
[3] Marmara University, Istanbul, Turkey

Abstract

A useful model for data analysis is partially nonlinear models (PNLRM) where response variable is represented as sum of a nonparametric and a parametric component. Since the model includes the coefficients of both the parametric and nonparametric parts, the complexity of the model will be high and its interpretation will be very difficult. In this study, we propose a procedure that not only achieves the sparseness but also smoothness for PNLRM to obtain a simpler model that better explains the relationship between the response and co-variates. In the proposed method, the fused lasso problem is taken into account where nonparametric components are expressed as a spline basis function, and then Fused Lasso estimation problem is built and expressed in terms of conic quadratic programming which is solved interior point method. An application study is conducted to evaluate the performance of the proposed method by considering some well-known performance measures. The results are compared against parametric nonlinear model.

Keywords

Nonlinear model, Nonparametric regression, Estimation, Fused lasso, Spline function, Continuous optimization.

References

[1] Ben-Tal, A. and Nemirovski, A. (2001). Lectures on Modern Convex Optimization: Analysis, Algorithms and Engineering Applications. *MPS-SIAM Series on Optimization, Philadelphia, SIAM.*

[2] Tibshirani, R., Saunders, M., Rosset, S., Ji Zhu, J. and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of Royal Statistical Society B 67(1)*, 91–108.

[3] Ruppert, D., Wand, M.P. and Carroll, R.J. (2003). *Semiparametric Regression.* Cambridge University Press.

[4] Hastie, T., Tibshirani, R. and Friedman, J.H. (2001). *The Elements of Statistical Learning.* New York, Springer.

Deep cascaded prediction model for picture fuzzy time series

Özge Cağcağ Yolcu and Ufuk Yolcu

Marmara University, Istanbul, Turkey

Abstract

Fuzzy sets and some derivatives, such as intuitionistic and picture fuzzy sets, have been used in the prediction literature. A time series whose observations are sets of fuzzy pictures creates a picture fuzzy-time series. This study presents a deep cascaded prediction model for picture fuzzy time series. In the time series prediction process, model construction is the transaction of estimating a function as a prediction tool. Such a predictor can be either an equation consisting of a linear regression or time series prediction model or an artificial neural network with a complex and non-linear structure. A linear predictor ignores non-linear relationships, while a non-linear predictor ignores linear relationships. However, almost all-time series contain both linear and non-linear relationships. From this point of view, in this study, a deep structured cascade forward neural network with multiple hidden layers is designed as a prediction tool and estimated in ways that produce superior predictions. The distinguishing characteristic of a cascade forward neural network is that each layer is connected to all previous layers. In addition, the use of sigmoid and linear activation functions in the hidden layers and the output layer, respectively, makes it superior to other neural networks. It can model linear and non-linear relationships between inputs and outputs together and simultaneously, thanks to this feature.
The predictive ability of the estimated deep cascaded model has been discussed over out-of-sample data sets for widely used time series in related literature via different error metrics and some illustrations. All findings obtained from analyses prove the outstanding performance of the proposed deep cascaded prediction model by comparison with state-of-the-art models.

Keywords

Picture fuzzy time series, Deep cascaded model, Cascade forward neural network, Prediction.

References

[1] Cagcag Yolcu, O., Bas, E., Egrioglu, E. and Yolcu, U. (2020). A new intuitionistic fuzzy functions approach based on hesitation margin for time-series prediction. *Soft Comput. 11*, 8211–8222.
[2] Egrioglu, E., Bas, E., Cagcag Yolcu, O. and Yolcu, U. (2019). Intuitionistic time series fuzzy inference system. *Eng. Appl. Artif. Intell. 82*, 175–183.
[3] Egrioglu, E., Bas, E., Yolcu, U. and Chen, M.Y. (2020). Picture fuzzy time series: Defining, modeling and creating a new forecasting method. *Eng. Appl. Artif. Intell. 88*, 103367.
[4] Fahlman, S.E. and Lebiere, C. (1992). The cascade-cor-relation learning architecture. In: *1992 SEG Annu. Meet.*, 1992.

MANOVA for doubly-exchangeable covariance structure

Ivan Žežula and Daniel Klein

P. J. Šafárik University, Košice, Slovakia

Abstract

We consider matrix-valued multivariate observation model with three-level doubly-exchangeable covariance structure. Using known two-sample test procedures, MANOVA tests are constructed, and exact distributions of the test statistics are derived. Possible methods of computing p-values and critical values of the distributions are suggested.

Keywords

Multivariate observations, Three-level data, Special variance structures, Mean testing, MANOVA.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the Contract No. APVV-17-0568.

References

[1] Žežula, I., Klein, D. and Roy, A. (2020). Mean Value Test for Three-Level Multivariate Observations with Doubly Exchangeable Covariance Structure. In: Holgersson, T., Singull, M. (Eds.), *Recent Developments in Multivariate and Random Matrix Analysis - Festschrift in Honour of Dietrich von Rosen* (pp. 335–349). Cham: Springer.

[2] Žežula, I., Klein D. and Roy, A. (2021). Testing of multivariate repeated measures data with block exchangeable covariance structure. In: Filipiak, K., Markiewicz, A., von Rosen, D. (Eds.), *Multivariate, Multilinear and Mixed Linear Models* (pp. 233–252). Cham: Springer.

Jordan algebra in estimation and testing hypotheses in multivariate normal models

Roman Zmyślony and Arkadiusz Kozioł

University of Zielona Góra, Poland

Abstract

In this presentation estimation of fixed and covariance parameters will be considered in multivariate normal models with special covariance structure i.e. block compound symmetric (BCS). The properties of the estimators will be analyzed. Under multivariate normality, the free-coordinate approach is used to obtain unbiased linear and quadratic estimates for the model parameters. Optimality of these estimators follows from sufficiency and completeness of their distributions. As unbiased estimators with minimal variance, they are consistent.
Also in presentation will be given new approach for testing hypotheses on the structure of covariance matrices in double multivariate data with block compound symmetric covariance structure. It is proved that ratio of positive and negative parts of best quadratic unbiased estimators (BQUE) provide an F-test for independence of blocks variables in double multivariate models. Simulation studies for comparison of powers between F-test and LRT will be presented.

Keywords

Best unbiased estimator, Block compound symmetric covariance structure, Double multivariate data, Positive and negative part of estimator, Structure of covariance matrices, Structure of mean vector, Testing hypotheses.

References

[1] Jordan, P., von Neumann, J. and Wigner, E. (1934). On an algebraic generalization of the quantum mechanical formalism. *The Annals of Mathematics 35(1)*, 29–64.
[2] Drygas, H. (1970). *The Coordinate-Free Approach to Gauss-Markov Estimation.* Berlin, Heidelberg: Springer.
[3] Johnson, R.A. and Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis, sixth ed.* Pearson Prentice Hall, Englewood Cliffs, NJ.
[4] Kozioł, A., Roy, A., Fonseca, M., Zmyślony, R. and Leiva, R. (2018). Free-coordinate estimation for doubly multivariate data. *Linear Algebra Appl. 547*, 217–239.
[5] Michalski, A. and Zmyślony, R. (1999). Testing hypotheses for linear functions of parameters in mixed linear models. *Tatra Mountains Mathematical Publications 17*, 103–110.
[6] Seely, J.F. (1971). Quadratic subspaces and completeness. *The Annals of Mathematical Statistics 42(2)*, 710–721.
[7] Zmyślony, R. (1978). A characterization of best linear unbiased estimators in the general linear model. *Lecture Notes in Statistics 2*, 365–373.

Part V

Posters

R shiny based interface for multiple testing

Juseong Park[1], Shinjune Kim[2], Youngjae Oh[3] and Jaesik Jeong[1]

[1] CNU - Cheonnam National University, Republic of Korea
[2] KREI - Korea Rural Economic Institute, Republic of Korea
[3] Statistics Korea, Republic of Korea

Abstract

When it comes to the discovery of biomarker gene, we need to find some genes which show statistically significant difference between two groups, i.e., disease v.s. normal group. In high-throughput data, since there are numerous genes (variables) which are obtained from each subject, biomarker discovery is categorized as multiple testing problem. Many multiple testing methods, which control false discovery rate (FDR) have been developed.
As a seminal work, Benjamini and Hochberg (BH) develop a method that use ordered p-values. They also show that the proposed method control FDR at the aimed level. Unlike BH, Efron and his colleagues formulate the multiple testing problem into the mixture framework. One dimensional test statistic is used in the two component mixture model. Ploner and his colleagues propose a new method which use two-dimensional test statistic. From another angle, Kim and his colleagues suggest another two-dimensional test statistic. Main difference is that the method test for each one-dimensional statistic separately and then combine both of them by using Bonferroni idea. Recently, Ramos and his colleagues develop a method, which is especially working for spiky null case.
We develop R shiny based interface, which include all multiple testing methods mentioned above.

Keywords

Mutiple testing, FDR (False Discovery Rate), R shiny.

References

[1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological) 57(1)*, 289–300.
[2] Efron, B., Tibshirani, R., D Storey, J. and Tusher, V. (2001). Empirical bayes analysis of a microarray experiment. *Journal of American statistical association 96(456)*, 1151–1160.
[3] Ploner, A., Calza, S., Gusnanto, A. and Pawitan, Y. (2006). *Bioinformatics 22(5)*, 556–565

Part VI

List of participants

Participants

1. **S. Ejaz Ahmed**
   Brock University, Ontario, Canada
2. **Stanislav Anatolyev**
   CERGE-EI, Charles University, Prague, Czech Republic
3. **Cristina Andrade**
   Instituto Politécnico de Tomar, Portugal
   Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal
4. **Patrícia Antunes**
   University of Beira Interior, Covilhã, Portugal
5. **Anthony C. Atkinson**
   Department of Statistics, London School of Economics, London, UK
6. **Oskar Maria Baksalary**
   Adam Mickiewicz University, Poznań, Poland
7. **Beste Hamiye Beyaztaş**
   Istanbul Medeniyet University, Istanbul, Turkey
8. **Ufuk Beyaztaş**
   Marmara University, Istanbul, Turkey
9. **Ana Cantarinha**
   Instituto Politécnico de Beja, Portugal
10. **Marinela Capanu**
    Memorial Sloan Kettering Cancer Center, New York, USA
11. **Francisco Carvalho**
    Instituto Politécnico de Tomar, Tomar, Portugal
12. **Mason Chen**
    Stanford OHS, USA
13. **Carlos A. Coelho**
    Nova University of Lisbon, Portugal
14. **Deliang Dai**
    Linnaeus University, Växjö, Sweden
15. **Cristina Dias**
    Instituto Politécnico de Portalegre, Portugal
    Universidade Nova de Lisboa, Portugal
16. **Efthymia Derezea**
    University of Kent, Canterbury, UK
17. **Susana Faria**
    University of Minho, Guimarães,, Portugal
18. **Dário Ferreira**
    University of Beira Interior, Covilhã, Portugal

19. **Sandra Maria Bargão Ferreira**
Department of Mathematics, Faculty of Sciences, University of Beira Interior, Covilhã, Portugal
20. **Katarzyna Filipiak**
Poznań University of Technology, Poznań, Poland
21. **Eva Fišerová**
Department of Mathematical Analysis and Applocations of Mathematics, Palacký University Olomouc, Olomouc, Czech Republic
22. **Miguel Fonseca**
Universidade Nova de Lisboa, Portugal
23. **Cinzia Franceschini**
Università degli Studi di Scienze Gastronomiche di Pollenzo, Italy
24. **Carla Francisco**
Universidade de Évora, Portugal
25. **Emelyne Umunoza Gasana**
Linköping University, Linköping, Sweden
26. **Massimiliano Giacalone**
University of Naples "Federico II", Naples, Italy
27. **Xiaomi Hu**
Wichita State University, Kansas, USA
28. **Malwina Janiszewska**
Poznań University of Life Sciences, Poznań, Poland
29. **Mateusz John**
Poznań University of Technology, Poznań, Poland
30. **Stanislav Katina**
Masaryk University, Brno, Czech Republic
Czech Academy of Sciences, Prague, Czech Republic
31. **Şevval Kılıçoğlu**
Atılım University, Ankara, Turkey
32. **Seunghyeon Kim**
Chonnam National University, Gwangju, Korea
33. **Daniel Klein**
P. J. Šafárik University in Košice, Košice, Slovakia
34. **Yuli Liang**
Linnaeus University, Växjö, Sweden
35. **Nicola Loperfido**
Università degli Studi di Urbino "Carlo Bo", Urbino, Italy
36. **Vitalii Makogin**
Ulm University, Germany
37. **Mahmoud Mansour**
Al-Azhar University, Nasr City, Cairo, Egypt

38. **Augustyn Markiewicz**
Poznań University of Life Sciences, Poznań, Poland
39. **Stepan Mazur**
Örebro University, Sweden
Linnaeus University, Växjö, Sweden
40. **João Tiago Mexia**
Universidade Nova de Lisboa, Portugal
41. **M. Cristina Miranda**
Universidade de Aveiro, Portugal
42. **Monika Mokrzycka**
Polish Academy of Sciences, Poznań, Poland
43. **Paulo Monteiro**
Universidade de Évora, Portugal
44. **Elsa Moreira**
Nova School of Sciences and Technology, Caparica, Portugal
45. **Mina Norouzirad**
Nova School of Science and Technology, Caparica, Portugal
46. **Célia Nunes**
Universiddade da Beira Interior, Covilhã, Portugal
47. **M. Manuela Oliveira**
Universidade de Évora, Portugal
48. **Teresa Oliveira**
CEAUL, Lisboa, Portugal
Universidade Aberta, DCeT, Portugal
49. **Victor De Oliveira**
The University of Texas at San Antonio, Texas, USA
50. **J.A. Pereira**
Universidade do Porto, Portugal
CEAUL, Lisboa, Portugal
Universidade Aberta, DCeT, Portugal
51. **Simo Puntanen**
Tampere University, Tampere, Finland
52. **Anabela Rocha**
Universidade de Aveiro, Portugal
53. **Paulo Canas Rodrigues**
Federal University of Bahia, Salvador, Brazil
54. **Dietrich von Rosen**
Swedish University of Agricultural Sciences, Uppsala, Sweden
Linköping University, Sweden
55. **Carla Santos**
Polytechnic Institute of Beja, Portugal
Nova University of Lisbon, Portugal

56. **Antonio Seijas-Macías**
Universidade da Coruña, Spain
Universidade de Lisboa, Portugal
57. **Han Lin Shang**
Macquarie University, Australia
58. **Maria-Zafeiria Spyropoulou**
University of Kent, Canterbury, UK
59. **Müjgan Tez**
Marmara University, Istanbul, Turkey
60. **Atik Touazi**
University of Algiers, Algeria
University of Bejaia, Algeria
61. **Denise Uwamariya**
Linköping University, Linköping, Sweden
62. **Xiaoqin Wang**
University of Gävle, Sweden
63. **Ilgim Yaman**
Giresun University, Turkey
64. **Fatma Yerlikaya-Özkurt**
Atılım University, Ankara, Turkey
65. **Li Yin**
Karolinska Institutet, Stockholm, Sweden
66. **Özge Cağcağ Yolcu**
Marmara University, Istanbul, Turkey
67. **Ivan Žežula**
P. J. Šafárik University in Košice, Košice, Slovakia
68. **Roman Zmyślony**
University of Zielona Góra, Zielona Góra, Poland

Index