PROGRAM AND ABSTRACTS

ADVANCES IN DIRECTIONAL STATISTICS (ADISTA14)

INTERNATIONAL WORKSHOP

http://adista14.ulb.ac.be/

FREE UNIVERSITY OF BRUSSELS (ULB) AND THE ATOMIUM

BRUSSELS, BELGIUM

20TH-22ND MAY, 2014

Organising Committee

Alvaro Cruces, SIDERO, Luxembourg

Pierre Jeurissen, Free University of Brussels (ULB), Belgium

Christophe Ley, Free University of Brussels (ULB), Belgium

Arthur Pewsey, University of Extremadura, Spain

Thomas Verdebout, Lille 3 University, France

Scientific Committee

Christophe Ley, Free University of Brussels (ULB), Belgium

Arthur Pewsey, University of Extremadura, Spain

Thomas Verdebout, Lille 3 University, France

Financial Sponsors

EQUIPPE, Lille 3 University, France

ECARES, Free University of Brussels (ULB), Belgium

WELCOME FROM THE ADISTA14 ORGANISING COMMITTEE

Dear Workshop Participants,

It is an enormous pleasure for us to be able to welcome you all to Brussels for the ADVANCES IN DIRECTIONAL STATISTICS (ADISTA14) Workshop. We are extremely grateful to you all for having taken up our invitation to attend.

The primary objectives of the workshop are to bring together workers in the field of Directional Statistics from all around the world to share their latest research findings, foster dialogue and future research collaborations between researchers, and identify interesting new problems and potential applications. We hope you will find the program of talks we have put together scientifically stimulating and that you will have plenty of opportunity to discuss research ideas with participants already known to you as well as those you will be meeting for the first time here at ADISTA14. When inviting speakers we attempted to achieve a balance across different areas of expertise as well as levels of research experience in the field; from researchers with long-established international reputations to students just embarking on their research careers in the field. We hope that your participation will greatly benefit from that mix of expertise and experience.

The first and last days of the workshop will take place at the European Center for Advanced Research in Economics and Statistics of the Free University of Brussels (Université libre de Bruxelles in French, and ULB for short). We would like to express our deep gratitude to the authorities of the ULB for allowing us to use its facilities. The second day of the workshop will be held in the Atomium which, with its structure made of connecting spheres and cylinders, seemed to us an ideal place to hold a workshop on Directional Statistics.

Brussels is the political capital of Europe and a major tourist destination, awash with culture and sights to see. We feel sure you will greatly enjoy the city of Brussels and your participation in the social events we have organised.

We would like to sincerely thank ECARES of the ULB and EQUIPPE of Lille 3 University, France, for their generous financial support of the workshop, and to Spain's SEIO for their institutional support.

We wish you a stimulating, productive and enjoyable ADISTA14 workshop, and a memorable stay in Brussels!

The ADISTA14 Organising Committee.

WORKSHOP PROGRAM

DAY 1. TUESDAY, 20TH MAY (ULB, ROOM R42.2.113)

08:00–08:45: Registration

08:45–09:00: Welcoming event

09:00–10:30: S1 *Miscellaneous Session 1* (Mardia, Kent, Ko)

10:30–11:00: Coffee break

11:00–12:30: S2 *Non-parametric and Smoothing Methods* (Di Marzio, Marinucci, Rodríguez-Casal)

12:30–14:00: Lunch at Séraphine

14:00–15:30: S3 *Miscellaneous Session 2* (Rivest, Kume, Garcia-Portugués)

15:30–19:00: Guided sightseeing tour of Brussels

DAY 2. WEDNESDAY, 21ST MAY (ATOMIUM)

09:30–11:00: S4 *Hypothesis Testing* (Jupp, Paindaveine, Ley)

11:00–11:30: Coffee break

11:30–12:30: S5 *Computer Intensive Methods* (Wood, Pewsey)

12:30–14:00: Lunch at the Atomium

14:00–15:30: S6 *Inference on Spheres* (Swan, Verdebout, Sabbah)

15:30–16:00: Coffee break

16:00–17:30: S7 *Flexible Models for Circular Statistics* (Jones, Kato, Abe)

20:00– : Conference dinner at Quincaillerie

DAY 3. THURSDAY, 22ND MAY (ULB, ROOM R42.2.113)

08:30–09:00: Coffee

09:00–10:30: S8 *Computational Statistics* (Taylor, Crujeiras, Agostinelli)

10:30–11:00: Coffee break

11:00–12:30: S9 *Applications* (Hamelryck, Theobald, Ehler)

12:30–14:00: Lunch at cafeteria

14:00–15:30: S10 *Time Series, Random Fields and Spatial Markov Models* (Gelfand, Jona-Lasinio, Lagona)

15:30–16:00: Coffee break

16:00–17:30: S11 *Miscellaneous Session 3* (Lockhart, Richter, SenGupta)

17:30– : Closing event

ABSTRACTS

Contents

S1 *Miscellaneous Session 1*

In-depth analysis of some shapes in proteins with applications: modelling conics and helices

Kanti Mardia
University of Oxford & University of Leeds

It is well-known that proteins are the workhorses of all living systems. So these are extremely important in drug discovery, medicine and evolutionary biology. However, there is a lot of mystery in their function; their malfunction leads to various diseases such as Alzheimer's, cancer and so on. The function depends on their three-dimensional atomic configuration (shape) which, interestingly, can be described in terms of certain sets of angles so directional statistics plays a key role. There are various levels at which the shape of proteins can be investigated. Here we will concentrate on what are called secondary structures: Helix, Beta-sheet (strands), and loops. In this context, the Ramachandran plot has now received a great deal of attention in directional statistics, but there are many other challenges needing statistics driven by geometry. Our main focus here is modelling a helix. This has become increasingly important as it is found that existence of a kink in a helix plays a central role in the function of so-called membrane proteins and thus in turn for drug discovery. We introduce a null model which has many ingredients. First, it incorporates the cylindrical nature of the helix. Next, for locating a kink, we use a change in direction of the axis (change-point extended to change-axis) of the helix. This can be reduced to the problem of finding the confidence interval between two mean directions from the Fisher distribution, and we examine the work of Lewis and Fisher on this problem. In fact, for a concentrated data, we obtain an elegant solution though our solution somewhat differs from theirs. Finally, we discuss many inference problems related to 'kink' versus 'no–kink'.

This is joint work with Charlotte Deane and Henry Wilman, Oxford University.

Modelling and simulation of directional distributions

John T. Kent
University of Leeds

Directional data analysis is concerned with statistical analysis on various non-Euclidean manifolds, starting with circle and the sphere, and extending to related manifolds such as Stiefel manifolds, Grassmann manifolds, rotation groups, shape spaces, and products of such manifolds. This talk will discuss (1) the construction of suitable statistical models on these spaces, generally based on exponential families, and (2) inference and simulation for these distributions. A recently developed method based on acceptance/rejection has greatly simplified the simulation task in a wide variety of situations.

SB-robust M-estimators for axial data

Daijin Ko

University of Texas at San Antonio

In earlier papers, Ko & Chang (1993) and Chang (2004) introduced SB robust M-estimators for location and concentration parameters for directional data and studied their properties. We extend the corresponding notions to axial data and/or great circle data and study statistical properties of M-estimators, including axial mean, axial median by Fisher, Lunn & Davies (1993), and the mle of the concentration parameter of Watson distribution. The standardized influence function and asymptotic distribution of these estimators are derived. Various senses of optimal robustness of the SB robust M-estimators for the axial data are discussed in light of He & Simpson (1993) and a simulation study is provided for finite sample properties of the estimators.

References

Ko, D. & Chang, T. (1993) Robust M-Estimators on Spheres. *Journal of Multivariate Analysis*, **45**, 104–136.

Chang, T. (2004) Spatial Statistics. *Statistical Science*, **19**, 624–635.

Fisher, N., Lunn, A. & Davies, S. (1993). Spherical median Axes. *Journal of the Royal Statistical Society: Series B*, **55**, 117–124.

He, X. & Simpson, D. (1992) Robust M-Estimators on Spheres. *Annals of Statistics*, **20**, 351–369.

S2 *Non-parametric and Smoothing Methods*

Local regression for circular data

Marco Di Marzio$^{(1)}$, Agnese Panzera$^{(2)}$ & Charles Taylor$^{(3)}$
University of Chieti-Pescara$^{(1)}$*, University of Florence*$^{(2)}$*, University of Leeds*$^{(3)}$

Local regression for circular data is a theme that has been developed only recently. We present methodological advances focusing on the kernel method, when the predictor and/or the response are circular. A number of applications are possible, from nonparametric estimation of trend in circular time series, to quantile estimation of circular distributions.

Normal approximations for linear and U-statistics on spherical Poisson fields

Solesne Bourguin$^{(1)}$, Claudio Durastanti$^{(2)}$, **Domenico Marinucci**$^{(2)}$ & Giovanni Peccati$^{(3)}$
Carnegie Mellon University$^{(1)}$*, University of Rome Tor Vergata*$^{(2)}$*, University of Luxembourg*$^{(3)}$

We review a recent stream of research on normal approximations for linear and U-statistics of wavelets/needlets coefficients evaluated on a homogeneous spherical Poisson field. We show how exploiting results from Peccati & Zheng (2011), based on Malliavin calculus and Stein's methods, it is possible to assess the rate of convergence to Gaussianity for a triangular array of statistics with growing dimensions. These results can be exploited in a number of statistical applications, such as spherical density estimations, searching for point sources, estimation of variance and the spherical two-sample problem.

Mode testing for circular data

Jose Ameijeiras-Alonso, Rosa M. Crujeiras & **Alberto Rodríguez-Casal**
University of Santiago de Compostela

Probability density functions fully describe the behavior of random variables, although in many cases, a complete characterization is not necessary and the main interest is to determine which variable values are most likely. For continuous random variables, this procedure is just the identification of the density modes (local maxima of the probability density function). There are several alternatives for detecting

modes in scalar variables. Although there are various exploratory tools, like the Sizer method proposed by Chaudhuri & Marron (1999), these do not provide a formal statistical hypothesis test. Within this context, it is worth mentioning the critical bandwidth, see Hall & York (2001), and also those tests based on the excess mass, see Müller & Sawitzki (1991). In this work, using the nonparametric kernel density estimator proposed in Di Marzio *et al.* (2009), these techniques are adapted to the circular data context and a new procedure for calibrating the excess mass statistics is presented. Its behavior is analyzed thought a extensive simulation study. Finally, the proposed method is used to study the seasonal patterns of forest fires in different geographic regions.

References

Chaudhuri, P. and Marron, J. S. (1999) SiZer for exploration of structures in curves. *Journal of the American Statistical Association*, **94**, 807–823.

Di Marzio, M., Panzera, A. & Taylor, C.C. (2009) Local polynomial regression for circular predictors. *Statistics & Probability Letters*, **79**, 2066–2075.

Hall, P. & York, M. (2001) On the calibration of Silverman's test for multimodality. *Statistica Sinica*, **11**, 515–536.

Müller, D.W. & Sawitzki, G. (1991) Excess mass estimates and tests for multimodality. *Journal of the American Statistical Association*, **86**, 738–746.

S3 *Miscellaneous Session 2*

Multivariate regression models for a dependent circular variable

Louis-Paul Rivest, Thierry Duchesne & Daniel Fortin
University of Laval

A general regression model for a dependent angular variable (y), with angular (x) and real (z) explanatory variables, is introduced. The predicted angle $\mu(y|x,z)$ is a compromise between the explanatory angles x whose relative importance might depend on the z variables; it generalizes the decentred predictor of Rivest (1997). The regression errors can be assumed to be independent and identically distributed; this gives the homogeneity model. They can also depend on the degree of agreement between the x variables. This defines the consensus model, similar to the proposal of Presnell *et al.* (1998). Maximum likelihood estimators for the parameters assuming that the error's distribution belongs to the von Mises family are proposed for the two models and robust sandwich variance estimates are derived.

This work is motivated by recent development in ecology where data on an animal's movement, collected using a GPS tracking device, is merged with data on its environment using a GIS (Geographic information system). In this context, one is interested in constructing a prediction model for y, the angle of an animal's movement at a given time, in terms of the directions x of targets of interest, such as its lair or a patch rich in nutriment, and the distances z to these targets. In ecology, these models are known as biased correlated random walks, see Benhamou (2014). The analysis of a data set to investigate how the direction of bison trails is influenced by the orientation of the next meadow and of the next canopy gap is presented as an illustration.

References

Benhamou, S. (2014) Of scales and stationarity in animal movements. *Ecology Letters*, **17**, 261–272.

Presnell, B., Morrison, S.P. & Littell, R.C. (1998) Projected multivariate linear models for directional data. *Journal of the American Statistical Association*, **93**, 1068–1077.

Rivest, L.P. (1997) A decentred predictor for circular-circular regression. *Biometrika*, **84**, 717–726.

On the explicit form of the Pfaffian of the Fisher-Bingham integral

Alfred Kume
University of Kent

Holonomic function theory is successfully implemented in a series of recent papers to efficiently calculate the normalizing constant and perform likelihood estimation for the Fisher-Bingham distributions. A key ingredient for establishing the standard holonomic gradient algorithms is the calculation of the Pfaffian equations. So far the present papers either calculate these symbolically or apply certain methods to simplify this process. Here we will show the explicit form of the Pfaffian equations using the expressions from Laplace inversion methods.

Assessing parametric regression models with directional predictors

Eduardo García–Portugués(1), Ingrid Van Keilegom(2), Rosa M. Crujeiras(1) & Wenceslao González–Manteiga(1)
University of Santiago de Compostela(1), Catholic University of Louvain(2)

A new test for assessing if the directional–linear regression function belongs to a certain parametric family is proposed in this work. The test is based on kernel smoothing and confronts the parametric hypothesis against a nonparametric alternative. The squared distance between the smoothed parametric estimate and a local linear estimator for the regression function, adapted to deal with a directional predictor, is used as test statistic. The asymptotic distribution of the test statistic is obtained, under simple and composite null hypotheses, and also for a family of local alternatives. A consistent resampling procedure for the practical calibration of the test is also provided. The performance of the method is illustrated in a simulation study. Finally, the test is applied to analyze several real datasets.

S4 *Hypothesis Testing*

Data-driven tests in directional statistics

Peter Jupp
University of St Andrews

In directional statistics the sample spaces are compact Riemannian manifolds. For any such manifold, M, there are canonical maps $\mathbf{t}_k$ $(k = 1, 2, \dots)$ of M into orthogonal finite-dimensional subspaces E_k of $L^2(M)$, the space of square-integrable functions.

Perhaps the hypotheses of greatest interest in directional statistics are (i) uniformity, (ii) symmetry under the action of a group, (iii) equality of several distributions, (iv) independence of random variables on manifolds M and N, (v) being in a given parametric model (goodness of fit). For each of these hypotheses, there are two appealing classes of tests based on the statistics $\mathbf{t}_{(k)} = \sum_{j=1}^{k} \mathbf{t}_j$: (a) those using score test statistics T_k, (b) those using simple 'non-standardised' forms S_k of T_k.

The choice of k is a matter of compromise: for small k, T_k and S_k can usually be evaluated easily; larger values of k give consistency against a wider class of alternatives. This talk will describe data-driven versions of the tests, in which the value $\hat{k}$ of k is chosen to maximise $U_k - \hat{E}(U_k)c(n)$, where U_k is T_k or S_k, n is sample size, c is a suitable increasing function, and $\hat{E}(U_j)$ is a consistent estimator of $E(U_j)$. The null hypothesis is rejected for large values of $U_{\hat{k}}$. Pleasant large-sample properties of these tests are (a) under the null hypothesis, $\hat{k}$ tends to be near 1, (b) under the alternative, $\hat{k}$ tends to be big enough to cause rejection of the null hypothesis.

Universal asymptotics for high-dimensional sign tests

Davy Paindaveine(1) & Thomas Verdebout(2)
Free University of Brussels (ULB)(1)*, Lille 3 University*(2)

We consider hypothesis testing for high-dimensional directional data. We mainly focus on the problem of testing uniformity on the unit sphere, and investigate the asymptotic behaviour of both following tests: the celebrated Rayleigh (1919) test and the Hallin & Paindaveine (2006) sign test. While we require both the sample size n and the dimension p to go to infinity, we allow these to do so in a totally arbitrary way, which is made possible by identifying a proper martingale structure in the sequences of test statistics at hand. This "universal" (n, p)-asymptotic investigation makes the above tests much more relevant in practice than the usual (n, p)-asymptotic one imposing that $p/n \to c$, for some c in a given convex set $C \subset (0, \infty)$. Simulations confirm our asymptotic results.

References

Hallin, M. & Paindaveine, D. (2006) Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Annals of Statistics*, **34**, 2707–2756.

Rayleigh, L. (1919) On the problem of random vibrations and random flights in one, two and three dimensions. *Philosophical Magazine*, **37**, 321–346.

Optimal tests for reflective/rotational symmetry

Christophe Ley(1) & Thomas Verdebout(2)

Free University of Brussels (ULB)(1)*, Lille 3 University*(2)

We propose optimal tests for circular reflective symmetry about a fixed median direction. The distributions against which optimality is achieved are the k-sine-skewed distributions of Umbach & Jammalamadaka (2009). We first show that sequences of k-sine-skewed models are locally and asymptotically normal in the vicinity of reflective symmetry. Following the Le Cam methodology, we then construct optimal (in the maximin sense) parametric tests for reflective symmetry, which we render semi-parametric by a studentization argument. These asymptotically distribution-free tests happen to be uniformly optimal (under any reference density) and are moreover of a very simple and intuitive form. They furthermore exhibit nice small sample properties, as we show through a Monte Carlo simulation study.

We finish the talk by extensions of this work in two directions: first, to testing for reflective symmetry about an unspecified center, and second to testing for rotational symmetry against skew-rotationally symmetric distributions that will be introduced in this talk.

References

Umbach, D. & Jammalamadaka, S.R. (2009) Building asymmetry into circular distributions. *Statistics and Probability Letters*, **79**, 659–663.

S5 *Computer Intensive Methods*

Regression Models for Directional Data

Andrew T.A. Wood
University of Nottingham

Various types of regression model have been considered for response variables which are unit vectors. The circle (2D) and sphere (3D) are the most important cases for applications. In an early paper, Gould (1969, Biometrics) considered regression models on the circle and sphere in which the angular variable in the circular case, and latitude and longitude of the mean direction in the spherical case, are modelled as linear combinations of the covariates. In subsequent work, a number regression models have been proposed for data on the circle and sphere, including: the regression model on the circle for a single linear covariate due to Johnson & Wehrly (1978, JASA); directional regression models in which both the response vector and the covariate vector are unit vectors of the same dimension (Chang, 1986 AoS; Chang, 1989 AoS; Rivest, 1989, AoS); the circular regression models of Fisher and Lee (1992, Biometrics); and circular (Downs & Mardia, 2002, Biometrika) and spherical (Downs, 2003, Biometrika) regression models based on types of Möbius transformation. Some, though not all, of the above models use the von Mises-Fisher distribution as the directional error distribution. Directional regression models based on the projected normal distribution have been considered by Presnell *et al.* (1998, JASA) and, from a Bayesian perspective, by Wang & Gelfand (2013, Statistical Methodology). In the talk I will consider regression models on the sphere in which it is not assumed that the directional error distribution possesses rotational symmetry. Both parametric and nonparametric approaches will be discussed.

Bootstrap goodness-of-fit testing for Wehrly-Johnson bivariate circular models

Arthur Pewsey
University of Extremadura

We consider goodness-of-fit testing for bivariate circular models generated using the popular general construction of Wehrly & Johnson (1980). Our approach hinges on, first, transforming the original pair of random variables to, if the model were correct, a pair distributed uniformly on the torus, and then applying existing tests for toroidal uniformity. Of course, in practice the parameters of any chosen model must be estimated. We use simulation to explore the size of such test procedures for data from the appealing bivariate wrapped Cauchy model of Kato & Pewsey (2013) when its parameters are estimated using maximum likelihood and various tests for toroidal uniformity of Wellner (1979) and Jupp (2009) are applied. The power of the test procedures against the bivariate von Mises distribution is also investigated.

References

Jupp, P.E. (2009) Data-driven tests of uniformity on product manifolds. *Journal of Statistical Planning and Inference*, **139**, 3820–3829.

Kato, S. & Pewsey, A. (2013) A Möbius transformation-induced distribution on the torus. *To appear*.

Wehrly, T. & Johnson, R.A. (1980) Bivariate models for dependence of angular observations and a related Markov process. *Biometrika*, **66**, 255–256.

Wellner, J.A. (1979) Permutation tests for directional data. *Annals of Statistics*, **7**, 929–943.

S6 *Inference on Spheres*

The Le Cam methodology for directional data

Christophe Ley[(1)], **Yvik Swan**[(2)], Baba Thiam[(3)] & Thomas Verdebout[(3)]
Free University of Brussels (ULB)[(1)], University of Liège[(2)], Lille 3 University[(3)]

We describe how to use the Le Cam methodology for estimation and testing problems for directional data. In particular we discuss the local asymptotic normality property of a sequence of rotationally symmetric models; this is a non-standard result due to the curved nature of the unit sphere. Locally and asymptotically most stringent parametric tests for ANOVA are constructed for directional data within the class of rotationally symmetric distributions. Asymptotic relative efficiencies are calculated and the finite-sample behaviour of the proposed tests is investigated using Monte Carlo simulation.

Inference for the concentration of rotationally symmetric distributions

Christophe Ley[(1)] & **Thomas Verdebout**[(2)]
Free University of Brussels (ULB)[(1)], Lille 3 University[(2)]

One-sample and multi-sample tests on the concentration parameter of rotationally symmetric distributions on (hyper-)spheres have been well studied in the literature. However, only little is known about their behaviour under local alternatives, which is due to complications inherent to the curved nature of the parameter space. The aim of the present paper therefore consists in filling that gap by having recourse to the Le Cam methodology. The finite-sample behaviour of the studied tests are investigated using Monte Carlo simulation.

Quantiles and depth for directional data

Christophe Ley[(1)], **Camille Sabbah**[(2)] & Thomas Verdebout[(2)]
Free University of Brussels (ULB)[(1)], Lille 3 University[(2)]

In univariate statistics, the properties of the quantiles of a distribution are well known. They provide information about both dispersion and central tendency. Moreover and because they are computable over an entire range of values, univariate quantiles naturally lead to comparison of distributions (QQ-plots) among various other applications and are widely used today.

Generalization of the notion of quantiles in the multivariate framework has proved to be a difficult task due to the absence of a natural order for vectors, and has been the subject of much attention in the recent years.

In the context of spherical data, we propose a new concept of depth (for unimodal distributions) which builds on the univariate quantile concept and inherits its asymptotic properties. More precisely, a Bahadur representation of these quantiles is given which turns out to be canonic for rotationally symmetric distributions. We conclude by giving several applications of our new concept of quantiles directly inspired from the ones available for univariate distributions (robust measures of concentration, QQ-plots, goodness-of-fit tests, trimming, etc.)

S7 *Flexible Models for Circular Statistics*

Families of Unimodal Distributions on the Circle

Chris Jones
The Open University

In something of an introduction to the talks of the other speakers in this session, I will start my talk on the real line, briefly describing some families of unimodal distributions with three or four parameters, controlling location, scale, skewness and perhaps some aspect(s) of tailweight. I will then address the question: "Can this technology be transferred to the case of distributions on the circle?" The answer is a qualified yes. Of the various possibilities that I look over, the one I will spend most time on is the "inverse Batschelet" distributions of myself and Arthur Pewsey (not unrelated, of course, to a suggestion of Batschelet!). However, this also forms a lead up to the following talk, by Shogo Kato, that explores a family of distributions emerging from this work which appears to prove best of all, possessing numerous attractive properties. Over to you, Shogo!

A tractable and interpretable four-parameter family of unimodal distributions on the circle

Shogo Kato(1) & M.C. Jones(2)
Institute of Statistical Mathematics(1)*, The Open University*(2)

On the circle, as on the line, families of unimodal distributions with parameters controlling location, scale or concentration, skewness and, in some appropriate sense, kurtosis, are useful for robust modelling. Although numerous such families now exist on the line, fewer exist on the circle.

In this talk we present a family of four-parameter distributions for circular data by taking a new approach. Properties of the proposed family include: unimodality; a simple characteristic function and tractable density and distribution functions; interpretable parameters individually measuring location, concentration, skewness and kurtosis, respectively; a wide range of skewness and "kurtosis"; some sub-models including the wrapped Cauchy and cardioid distributions; closure under convolution and multiplication by certain constants; straightforward parameter estimation by both method of moments (suitable for smaller samples and moderate parameter values) and maximum likelihood. We will show that our new proposal compares favourably with some of the current four-parameter unimodal families on the circle. An illustrative application of the proposed model is given.

We stress unimodality because bi- and multi-modal distributions can most interpretably be modelled by mixtures of unimodal components or perhaps, on the circle, by 'multiplicative mixtures' of which the distributions of interest may form components.

Skewed circular distributions with unimodality and mode-preserving property

Toshihiro Abe[1], Arthur Pewsey[2] & Hironori Fujisawa[3]

Tokyo University of Science[1], *University of Extremadura*[2], *Institute of Statistical Mathematics*[3]

We consider unimodal skew-symmetric circular distributions derived using a monotone inverse function. General results for the distributions are presented, including distribution and density based circular skewness measures. The family of inverse k-sine-skewed circular distributions is introduced as a special case. More flat-topped and sharply peaked versions of the distributions are also considered. General results are provided for maximum likelihood estimation, including the Fisher information matrix.

S8 *Computational Statistics*

Some computational issues in nonparametric circular-circular regression

Charles Taylor(1), Marco Di Marzio(2) & Agnese Panzera(3)
University of Leeds(1)*, University of Chieti-Pescara*(2)*, University of Florence*(3)

We consider rotation models which are solutions obtained from weighted least squares, the weights being a function of circular distance depending on a smoothing parameter. Adaptations for the case in which the mean response is a reflection, rather than a rotation are investigated. By considering terms in a series expansion, a "local linear" model is proposed (for the case of circular data) and extensions to higher-dimension data discussed.

Assessing significant features in nonparametric circular curves

María Oliveira(1), **Rosa M. Crujeiras**(2) & Alberto Rodríguez–Casal(2)
Durham University(1)*, University of Santiago de Compostela*(2)

In nonparametric circular density and circular–linear regression estimation, the smoothing parameter controls the global appearance of the estimator and its dependence on the sample, in such a way that an unsuitable choice of this value may provide a misleading estimate of the density or the regression curve. Hence, the assessment of the statistical significance of the observed features through the smoothed curve is required for not compromising the extracted conclusions. An approach to circumvent the choice of the smoothing parameter, and still be able to assess global structure features in the curve, is given by the SiZer method developed by Chaudhuri & Marron (1999) for the analysis of linear data. An extension of the original SiZer to circular data, namely CircSiZer, has been introduced by Oliveira *et al.* (2014).

For illustrating the idea of CircSiZer, consider a random sample of size 250 from model M10 from Oliveira *et al.* (2012). Figure 8.1 (left panel) shows the family of nonparametric smoothers for a wide range of values of the smoothing parameter. CircSiZer focuses on finding regions where the gray curves are significantly increasing/decreasing (blue/red) and displaying this information in a color map as the one shown in Figure 8.1 (right panel). Different ways for constructing these regions, based on confidence intervals for the derivative, both for circular density and circular–linear regression will be discussed in this talk.

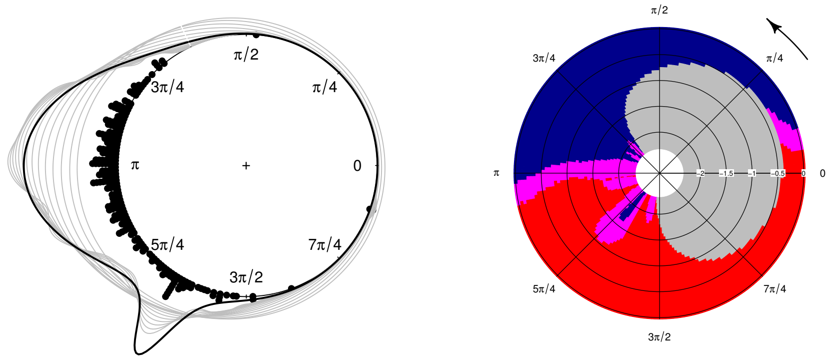

Figure 8.1: Left panel: family of kernel density estimates (gray curves) indexed by the smoothing parameter for a sample of 250 data. Right panel: CircSiZer map for kernel density estimator.

References

Chaudhuri, P. & Marron, J.S. (1999) SiZer for exploration of structures in curves. *Journal of the American Statistical Association*, **94**, 807–823.

Oliveira, M., Crujeiras, R.M. & Rodríguez–Casal, A. (2012) A plug–in rule for bandwidth selection in circular density estimation. *Computational Statistics and Data Analysis*, **56**, 3898–3908.

Oliveira, M., Crujeiras, R.M. & Rodríguez–Casal, A. (2014) CircSiZer: an exploratory tool for circular data. *Environmental and Ecological Statistics*, **21**, 143–159.

Circular: an R package for circular data analysis

Claudio Agostinelli

Ca' Foscari University of Venice

Circular data appear in different scientific fields like in Natural, Physical, Medical and Social Sciences in which some or all the measurements are directions. A circular observation can be regarded as a point on a circle of unit radius, or a unit vector (i.e. a direction in the plane). Once an initial direction and orientation of the circle have been chosen, each circular observation can be specified by the angle from the initial direction to the point on the circle corresponding to the observation. Because of the nature of circular observations, their statistical analysis (descriptive and/or inferential) cannot be carried out using standard methods used for observations in Euclidean space. For this reason, the statistical analysis of circular data needs specific methods and software. The R package `circular` implements a unified framework to manipulate, visualize and analyze circular data. We will show how to perform robust statistical analysis for circular data modelled using the von Mises or the wrapped normal distributions using functions available in the R package `wle`. Finally, we illustrate a nonparametric analysis based on (local) angular simplicial depth and (local) angular Tukey depth using functions available in the R package `localdepth`.

S9 *Applications*

Probabilistic models of protein structure in atomic detail and their applications

Thomas Hamelryck
University of Copenhagen

The so-called protein folding problem is the loose designation for an amalgam of closely related, unsolved problems that include protein structure prediction, protein design and the simulation of the protein folding process (Hamelryck *et al.*, 2012). We adopt a unique probabilistic approach to modeling bio-molecular structure, based on graphical models, directional statistics (Boosma *et al.*, 2008; Harder *et al.*, 2010) and probability kinematics (Hamelryck *et al.*, 2010; Hamelryck *et al.*, 2013). Notably, we developed a generative probabilistic model of protein structure in full atomic detail (Valentin *et al.*, 2013). I will give an overview of how rigorous probabilistic models of something as complicated as a protein's atomic structure can be formulated, focusing on the use of graphical models and directional statistics to model angular degrees of freedom. I will also discuss the reference ratio method, which is needed to "glue" several probabilistic models of protein structure together in a consistent way. The reference ratio method is based on "probability kinematics", a little known alternative to Bayesian updating proposed by the philosopher Richard C. Jeffrey at the end of the 50's. Probability kinematics might find widespread application in statistics and machine learning as a way to formulate complex, high-dimensional probabilistic models by combining several simple models.

References

Boomsma, W., Mardia, K.V., Taylor, C.C., Ferkinghoff-Borg, J., Krogh, A. & Hamelryck, T. (2008) A generative, probabilistic model of local protein structure. *Proceedings of the National Academy of Science of the USA*, **105**, 8932–8937.

Harder, T., Boomsma, W., Paluszewski, M., Frellsen, J., Johansson, K.E. & Hamelryck, T. (2010) Beyond rotamers: a generative, probabilistic model of side chains in proteins. *BMC Bioinformatics*, **11**, 306.

Hamelryck, T., Borg, M., Paluszewski, M., Paulsen, J., Frellsen, J., Andreetta, C., Boomsma, W. Bottaro, S. & Ferkinghoff-Borg, J. (2010) Potentials of mean force for protein structure prediction vindicated, formalized and generalized. *PLoS ONE*, **5**, e13714.

Hamelryck, T., Mardia, K.V. & Ferkinghoff-Borg, J. (Editors) (2012) *Bayesian Methods in Structural Bioinformatics*. New York: Springer Verlag.

Hamelryck, T., Haslett, J., Mardia, K.V., Kent, J.T., Valentin, J., Frellsen, J. & Ferkinghoff-Borg, J. (2013) On the reference ratio method and its application to statistical protein structure prediction. In *LASR 2013 - Statistical Models and Methods for Non-Euclidean Data with Current Scientific Applications*. Leeds: Leeds University Press, pp. 53–57.

Valentin, J., Andreetta, C., Boomsma, W., Bottaro, S., Ferkinghoff-Borg, J., Frellsen, J., Mardia, K.V., Tian, P. & Hamelryck, T. (2013) Formulation of probabilistic models of protein structure in atomic detail using the reference ratio method. *Proteins*, **82**, 288–299.

Bayesian superpositions using MCMC sampling

Douglas L. Theobald[(1)], Richard Roy[(1)] & Kanti V. Mardia[(2)]
Brandeis University[(1)], *University of Leeds*[(2)]

The shape of an object is the geometrical information that is invariant to translation, rotation, and scale transformations. To analyze the shape differences in direct Euclidean space, one must first superpose them in some optimal fashion. Superposition thus involves inferring the optimal similarity transformations. We encode an object as a $k \times d$ matrix (a configuration of k Cartesian landmarks in d dimensions). In biophysics, for instance, the structures of macromolecules are described concisely by three-dimensional atomic coordinates. Macromolecular structures are conventionally superposed using least-squares, which assumes that landmarks are uncorrelated and have equal variance. In previous work, we reformulated the superposition problem with a probabilistic model in which structures are distributed normally according to an arbitrary covariance matrix (Theobald & Wultke, 2006; Theobald & Steindel, 2012). Identifiability problems were overcome using a conjugate inverse Wishart hyperprior for the covariance matrix. An expectation-maximization algorithm that finds the maximum likelihood superposition has been implemented in our program THESEUS, which is widely used in the structural bioinformatics community (Theobald & Wuttke, 2006; Theobald & Steindel, 2012). We have recently explored a full Bayesian extension of the matrix normal multiple superposition model that allows for scaling (Mardia *et al.*, 2013; Theobald & Mardia, 2011). The conditional posterior distribution of the scale factors has a non-standard form, and their use leads to new identifiability issues involving pathological correlations among the rotation, covariance, and scale parameters. Here we describe our updated THESEUS program, which determines the posterior distribution of all superposition parameters using MCMC and a novel acceptance-rejection method for sampling the 3D orthogonal rotations (Kent *et al.*, 2013).

References

Kent, J.T., Ganeiber, A.M. & Mardia, K.V. (2013) A new method to simulate the Bingham and related distributions in directional data analysis with applications. http://arxiv.org/abs/1310.8110.

Mardia, K.V, Fallaize, C.J., Barber, S., Jackson, R.M. & Theobald, D.L. (2013) Bayesian alignment of similarity shapes. *Annals of Applied Statistics*, **7**, 989–1009.

Theobald, D.L. & Wuttke, D.S. (2006) Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian Procrustes problem. *Proceedings of the National Academy of Science of the USA*, **103**, 18521–18527.

Theobald, D.L. & Wuttke, D.S. (2006) THESEUS: Maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics*, **22**, 2171–2172.

Theobald, D.L. & Mardia, K.V. (2011) Full Bayesian analysis of the generalized non-isotropic Procrustes problem with scaling. In *Next Generation Statistics in Biosciences. Proceedings of the 30th Leeds Annual Statistical Research (LASR) Workshop* (Editors: A. Gusnanto, K.V. Mardia & C.J. Fallaize). Leeds: Department of Statistics, pp. 41–44.

Theobald, D.L. & Steindel, P.A. (2012) Optimal simultaneous superpositioning of multiple structures with missing data. *Bioinformatics*, **28**, 1972–1979.

From directional statistics to applications in signal processing

Martin Ehler

University of Vienna

Frames generalize the basis concept in vector spaces and have proven useful in fields like signal processing and compressed sensing. We merge directional statistics with frame theory and use those results in two ways, first to analyze patterns found in granular rod experiments and secondly to enhance filter banks in signal processing.

S10 *Time Series, Random Fields and Spatial Markov Models*

Analyzing spatial and spatio-temporal angular and linear data using Gaussian processes

Alan E. Gelfand
Duke University

Circular data arise in oceanography (wave directions) and meteorology (wind directions), and, also, with periodic measurements scaled to a circle. In this talk we introduce a hierarchical Bayesian framework to handle directional data taken at spatial locations, anticipating structured dependence between these measurements. Models are fitted straightforwardly using Markov chain Monte Carlo, introducing latent variables. Full inference, including uncertainty, emerges.

First, we formulate a wrapped Gaussian spatial process model for this setting, induced from a customary *inline* Gaussian process. Then, we consider the projected Gaussian spatial process built from a *bivariate* Gaussian process model. We look at the properties of these processes, including the induced correlation structure. We show that they can immediately be extended to accommodate spatio-temporal angular data. Our approach enables spatial interpolation/kriging and can accommodate measurement error. We illustrate with a set of spatio-temporal wave direction data from the Adriatic coast of Italy, generated through a complex computer model. The computer model also generates wave heights. This leads to construction of joint spatial and spatio-temporal models for wave height and wave direction.

Wrapped Gaussian processes: a short review and some new results

Giovanna Jona-Lasinio(1), Gianluca Mastrantonio(2) & Alan E. Gelfand(3)
Sapienza University of Rome(1), Rome 3 University(2), Duke University(3)

In the present work we review modeling strategies based on wrapped Gaussian processes defined to model spatial and spatio-temporal data. We first illustrate the model-based approach to handle periodic data in the case of measurements taken at spatial locations, anticipating structured dependence between these measurements. The wrapped Gaussian spatial process is here induced by a customary linear Gaussian process. We formulate the model as a Bayesian hierarchical one and we show that the fitting of the model is possible using standard Markov chain Monte Carlo methods. Then we move to some spatio-temporal generalizations of the above model. We present simulations studies of these proposals aiming at understanding their computational and statistical properties. We address the modeling of wave directions produced by a complex computer model of the Adriatic sea under various settings. We highlight the pros and cons of each model and the difficulties arising for the implementations of the MCMCs. Eventually we provide some general advice on the use of the spatio-temporal wrapped Gaussian process.

A hidden Markov approach to the analysis of space-time environmental data with linear and circular components.

Francesco Lagona
Rome 3 University

The analysis of space-time series with linear and circular components is complicated by multiple sources of correlation, which arise across time and space and between variables, and by the different supports on which the variables are observed; the real line and the circle. I describe a multivariate hidden Markov model that parsimoniously includes these features of the data within a single framework. The model integrates a circular von Mises Markov field and a Gaussian Markov field, with parameters that evolve in time according to a latent Markov chain. It permits a description of the data by means of a finite number of time-varying latent regimes, associated with easily interpretable components of large-scale and small-scale spatial variation of mixed linear and circular data. I also illustrate a computationally feasible Expectation-Maximization algorithm that allows for efficient maximum likelihood estimation of the parameters. I finally discuss a case study of sea currents in the Northern Adriatic sea, where the model provides a parsimonious representation of the sea surface by means of alternating environmental states.

S11 *Miscellaneous Session 3*

Bayes assisted goodness-of-fit for von Mises regression

Richard Lockhart, Michael Stephens & Zheng Sun
Simon Fraser University

For responses y_i on the unit circle in R^2 to predictors $x_i \in R^p$ we consider goodness-of-fit for the von Mises regression model

$$y_i|x_i \sim f(y_i|x_i) = \frac{\exp\{\theta_i^T y_i\}}{2\pi I_0(\|\theta_i\|)}$$

where

$$\theta_i^T = x_i\beta.$$

Here $\theta_i \in R^2$, x_i is $1 \times p$, β is $p \times 2$. Our focus is on testing the von Mises assumption. We apply a sample size dependent prior to the likelihood ratio of an alternative density to a density in the null hypothesis. For a given prior on the null, that is, on β, we derive an approximate test statistic which maximizes the power averaged over our priors. Adjusting the prior produces some well known goodness-of-fit statistics; the prior for which a test statistic is optimal is then informative about the qualities of the test statistic. I will highlight the structure of the resulting test statistic and consider simplifications which arise when, as in our von Mises regression model, there is a low dimensional sufficient statistic for the problem.

Polar angle distributions on generalized circles

Wolf-Dieter Richter
University of Rostock

It is well known that the von Mises density may be considered as the conditional density of the random polar angle given the Euclidean norm $R = \sqrt{X^2 + Y^2}$ of a two-dimensional Gaussian random vector $(X, Y)^T$ which is centered away from the origin. This approach to an angular distribution will be extended here under the more general assumption that the random vector follows an arbitrary norm contoured or even a star-shaped distribution which is symmetric with respect to its center. The suitably shifted density level sets of such a distribution are the boundaries of the scalar multiples of a convex body or of a star body, K, having its center at the origin, respectively. Moreover, the conditioning variable R will be changed with the value $h_K(X, Y)$ of the Minkowski functional h_K of the star body K, taken at the random point $(X, Y)^T$. The star body K is just an Euclidean ball, centered at the origin, if the random vector $(X, Y)^T$ is a shifted standard Gaussian or a spherical one, and its possible form can be adapted in a flexible way to the form of the cloud of points corresponding to a given data set in R^2. The form of K may alternatively satisfy some otherwise motivated probabilistic assumptions. The tails of the probability distribution of the random vector $(X, Y)^T$ may be chosen according to the situation under consideration both as heavy or light ones. Following this line, we present and exploit new geometric and stochastic representations generalizing those known for elliptically contoured distributions from Richter (2013) and for $l_{n,p}$-symmetric distributions from Richter (2009). From a technical point of view, we make use of the generalized trigonometric functions and generalized polar coordinates introduced in Richter (2011). For a discussion of certain properties of different types of angular and radial coordinates of Gaussian vectors, see Dietrich *et al.* (2013).

References

Dietrich, T., Kalke, S. & Richter, W.-D. (2013) Stochastic representations and a geometric parametrization of the two-dimensional Gaussian law. *Chilean Journal of Statistics*, **4**, 27–59.

Richter, W.-D. (2009) Continuous $l_{n,p}$-symmetric distributions. *Lithuanian Mathematical Journal*, **49**, 93–108.

Richter, W.-D. (2011) Circle numbers for star discs. *ISRN Geometry*, Article 479262.

Richter, W.-D. (2013) Geometric and stochastic representations for elliptically contoured distributions. *Communications in Statistics - Theory & Methods*, **42**, 579–602.

On some constructions and inference problems for spherical distributions

Ashis SenGupta

University of California-Riverside & Indian Statistical Institute

Some methods of constructing distributions on the unit sphere are presented, including one based on stereographic projection which yields a variant of a bi-modal t-distribution on the sphere. These methods also yield variants of normal, t, Cauchy, etc. distributions on the unit sphere. Some results for decision-theoretic estimation, e.g. admissibility, minimaxity, etc., of the simultaneous MLE of the mean direction vector parameters of several independent Fisher-Langevin distributions are also presented. Model-based cluster analysis using the Fisher-Langevin distribution is enhanced. Change-point analysis for the mean direction of this distribution using a Bayesian approach is discussed. Several real-life examples are presented to illustrate the methods developed.