Mathematisches Forschungsinstitut Oberwolfach

Report No. 28/2019

DOI: 10.4171/OWR/2019/28

Statistical Methodology and Theory for Functional and Topological Data

Organized by
Aurore Delaigle, Melbourne
Alexander Meister, Rostock
Victor Panaretos, Lausanne
Larry Wasserman, Pittsburgh

16 June – 22 June 2019

Abstract. The workshop focuses on the statistical analysis of complex data which cannot be represented as realizations of finite-dimensional random vectors. An example of such data are functional data. They arise in a variety of climate, biological, medical, physical and engineering problems, where the observations can be represented by curves and surfaces. Fast advances in technology continuously produce a deluge of bigger data with even more complex structures such as arteries in the brain, bones of a human body or social networks. This has sparked enormous interest in more general statistical problems where the random observations are elements of abstract topological spaces.

The workshop will stimulate development of new statistical methods for these types of data and will be an ideal platform for discussing their theoretical properties (e.g. asymptotic optimality), computational performance, and new exciting applications in areas such as machine learning, image analysis, biometrics and econometrics.

Mathematics Subject Classification (2010): 62Mxx, 62Hxx.

Introduction by the Organizers

The (half-size) Oberwolfach Workshop *Statistical Methodology and Theory for Functional and Topological Data* (1925b), which was organized by Aurore Delaigle (Melbourne, Australia), Alexander Meister (Rostock, Germany), Victor Panaretos (Lausanne, Switzerland) and Larry Wasserman (Pittsburgh, USA), was attended by about 25 participants from Australia, France, Germany, Hungary, Singapore, Spain, Switzerland, the UK and the USA. The main concept of the conference was to advance and promote the research for the analysis of complex data, which cannot be described as realizations of random vectors in a finite-dimensional Euclidean space. In particular the focus was on two settings: the analysis of observed (or partially observed) random functions and the analysis of observations in quite general topological spaces. There is also overlap between these branches of statistics with respect to both the scientific topics and the researchers who are involved. On the other hand the workshop contributed to the interaction between both areas. There were talks with emphasis on diverse types of application e.g. to word data from documents, recorded curves indicating emotions, flight curves of airplanes etc.. Other talks focused on theoretical/mathematical issues such as topological invariants of random objects, asymptotic theory for principal components, functional data on manifolds, probabilistic analysis of empirical Fréchet means etc.. We think that all talks brought very interesting and fruitful insights to the audience. They showed a lot of academic progress in the understanding of the statistical analysis for non-standard types of data in recent years; but also that a lot of work remains to be done in order to attain a solid and complete statistical framework for the analysis of new types of data, whose complexity is supposed to increase further in future by computational progress in the process of recording data. The organizers would like to thank all participants and the MFO administration for all effort and, also, for the support of a participant by the SVP program.

Acknowledgement: The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-1641185, "US Junior Oberwolfach Fellows". Moreover, the MFO and the workshop organizers would like to thank the Simons Foundation for supporting Katharine Turner in the "Simons Visiting Professors" program at the MFO.

Workshop: Statistical Methodology and Theory for Functional and Topological Data

Table of Contents

Abstracts

Total Variation Regularized Fréchet Regression for Metric-Space Valued Data

Hans-Georg Müller
(joint work with Zhenhua Lin)

Non-Euclidean data that are indexed with a scalar predictor such as time are increasingly encountered in data applications, while statistical methodology and theory for such random objects are not well developed yet. Random objects ([9]) have also been referred to as Object-Oriented Data ([18, 8]). Since basic notions introduced by Fréchet [3] typically play a major role, the statistical analysis of samples of random objects could be characterized as Fréchet analysis.

A specific challenge when dealing with random objects is the absence of linear operations. Typical examples where random objects are encountered include random samples of densities or more generally, probability measures; random covariance matrices and surfaces; samples of networks and trees. An important component of statistical analysis for random objects is the choice of the metric that defines the metric space, where the random objects are situated.

An important example is provided by random densities, for which the Wasserstein metric has been shown to work well in applications for one-dimensional densities. From a more theoretical perspective, this metric has become popular due to its connections with optimal transport ([11]). As a consequence, population and sample Fréchet-Wasserstein means, also known as barycenters, have emerged as useful statistical summaries of densities. For special subclasses of random objects, local linearization is a convenient tool that is commonly employed when the random objects are located on smooth manifolds ([1, 19, 10]). In other cases one may employ transformations to a linear space ([13]).

The inclusion of covariate information, a central tenet of statistics, provides motivation to go beyond Fréchet means. The concept of Fréchet regression ([14]), which is a general approach to regression when responses are complex random objects in a metric space and predictors are in $\mathcal{R}^p$, is based on the idea of extending the classical concept of a Fréchet mean to the notion of a conditional Fréchet mean. Generalized versions of both global least squares regression and local weighted least squares smoothing have been developed in terms of both theory and applications ([12]), extending previous approaches ([16, 2]). The target quantities are appropriately defined population versions of global and local regression for random object responses. Asymptotic rates of convergence for the corresponding fitted regressions using observed data to the population can be derived by applying empirical process methods for M estimators.

A promising regularization approach for nonparametric Fréchet regression with metric-space valued response variables and a one-dimensional scalar predictor variable aims at minimizing the sum of squared distances between targets and fitted values under a total variation based penalty of the fitted function, by employing an appropriate modification of the definition of total variation that covers metric-space valued functions ([6]). This approach builds on the rich literature for penalized least squares with total variation penalty, which is a popular method in signal processing ([15, 7, 4, 5, 17]). We show that the total variation regularized Fréchet estimator leads to a metric-space valued step function. The class of step functions is not only sufficiently powerful to approximate any function of finite total variation, but also advantageous in modeling functions that are discontinuous, since it automatically incorporates jumps.

For the case of random objects in Hadamard spaces we provide a detailed asymptotic analysis for the proposed regularized Fréchet regression. To obtain these results, we first establish some properties of pseudo-inner products in such spaces. This geometric analysis then leads to results on asymptotic minimax rates of convergence for the function estimates. When the target function is a step function in Hadamard space, these results can be applied to obtain convergence results for the estimates of location and size of the jump points.

Research supported by NSF and NIH grants.

References

[1] Bhattacharya, R. and Patrangenaru, V. (2005) Large sample theory of intrinsic and extrinsic sample means on manifolds-II. *Annals of Statistics*, **33**, 1225–1259.

[2] Faraway, J. J. (2014) Regression for non-Euclidean data using distance matrices. *Journal of Applied Statistics*, **41**, 2342–2357.

[3] Fréchet, M. (1948) Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, **10**, 215–310.

[4] Kim, S.-J., Koh, K., Boyd, S. and Gorinevsky, D. (2009) ℓ_1 trend filtering. *SIAM Review*, **51**, 339–360.

[5] Lellmann, J., Strekalovskiy, E., Koetter, S. and Cremers, D. (2013) Total variation regularization for functions with values in a manifold. In *2013 IEEE International Conference on Computer Vision*, 2944–2951. IEEE.

[6] Lin, Z. and Müller, H.-G. (2019) Total variation regularized Fréchet regression for metric-space valued data. *arXiv preprint arXiv:1904.09647*.

[7] Mammen, E. and van de Geer, S. (1997) Locally adaptive regression splines. *Annals of Statistics*, **25**, 387–413.

[8] Marron, J. S. and Alonso, A. M. (2014) Overview of object oriented data analysis. *Biometrical Journal*, **56**, 732–753.

[9] Müller, H.-G. (2016) Peter Hall, Functional Data Analysis and Random Objects. *Annals of Statistics*, **44**, 1867–1887.

[10] Panaretos, V. M. and Zemel, Y. (2016) Amplitude and phase variation of point processes. *The Annals of Statistics*, **44**, 771–812.

[11] Panaretos, V. M. and Zemel, Y. (2019) Statistical aspects of Wasserstein distances. *Annual Review of Statistics and its Application*, **6**, 405–431.

[12] Petersen, A., Deoni, S. and Müller, H.-G. (2019) Fréchet estimation of time-varying covariance matrices from sparse data, with application to the regional co-evolution of myelination in the developing brain. *The Annals of Applied Statistics*, **13**, 393–419.

[13] Petersen, A. and Müller, H.-G. (2016) Functional data analysis for density functions by transformation to a Hilbert space. *Annals of Statistics*, **44**, 183–218.

[14] Petersen, A. and Müller, H.-G. (2019) Fréchet regression for random objects with Euclidean predictors. *The Annals of Statistics*, **47**, 691–719.

[15] Rudin, L. I., Osher, S. and Fatemi, E. (1992) Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, **60**, 259–268.

[16] Steinke, F., Hein, M. and Schölkopf, B. (2010) Nonparametric regression between general Riemannian manifolds. *SIAM Journal on Imaging Sciences*, **3**, 527–563.

[17] Tibshirani, R. J. (2014) Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, **42**, 285–323.

[18] Wang, H. and Marron, J. S. (2007) Object oriented data analysis: Sets of trees. *Annals of Statistics*, **35**, 1849–1873.

[19] Yuan, Y., Zhu, H., Lin, W. and Marron, J. S. (2012) Local polynomial regression for symmetric positive definite matrices. *Journal of Royal Statistical Society, Series B*, **74**, 697–719.

Mean and Covariance Estimation for Functional Snippets

Jane-Ling Wang
(joint work with Zhenhua Lin)

Functional data are random functions on a common domain, e.g., an interval $\mathcal{T} \subset \mathbb{R}$. In reality they can only be observed on a discrete schedule, possibly intermittently, which leads to an incomplete data problem. Luckily, by now this problem has largely been resolved ([11, 7, 13]) and there is a large literature on the analysis of functional data. In this paper, we consider estimation of the mean and the covariance functions of functional snippets, which are short segments of functions possibly observed irregularly on an individual specific subinterval that is much shorter than the entire study interval. Such data occur frequently in longitudinal studies when subjects enter the study at random time and are followed for a short period within the domain $\mathcal{T} \subset \mathbb{R}$. For illustration purpose, we assume that $\mathcal{T}$ is the unit interval $[0, 1]$ and each subject is followed for a period of length δ that is much smaller than 1. As a result, the design of support points where one has information about the covariance function $\mathcal{C}(s, t)$ is incomplete in the sense that there are no design points in the off-diagonal region, $\{(s, t) : |s - t| > \delta, 0 \leq s, t \leq 1\}$. One therefore does not have information to locally estimate the covariance in such a region.

Functional snippets have come under different names as censored functional data ([2]), fragmented functional data ([3]), functional fragments ([4]), or partially observed functional data ([5]). These terminologies are used interchangeably with another type of partially observed data ([6, 9, 10, 8]), for which the span of a single individual curve can be as large as the span of the study. However, estimation of the covariance function for functional snippets is more challenging since information for the far off-diagonal regions of the covariance structure is completely missing. To avoid confusion and borrowing from the term longitudinal snippets in [1], we adopt the term "functional snippet", which distinguishes functional snippets from other partially observed functional data.

Previous works on functional snippets include [2, 3, 1, 4, 12, 5]. We took a different approach by addressing the difficulty of covariance estimation through decomposing the covariance function into a variance function component and a correlation function component. The variance function can be effectively estimated nonparametrically using data within the diagonal band $\{(s,t) : |s-t| \leq \delta, 0 \leq s,t \leq 1\}$, while the correlation part is modeled parametrically so missing information in the far off-diagonal regions can be extrapolated from data within the diagonal band. Both theoretical analysis and numerical simulations suggest that this hybrid strategy is effective and efficient. Our theory also allows increasing number of parameters, thus extending the semi-parametric hybrid approach to a nearly nonparametric paradigm. In addition, we propose a new estimator for the variance of measurement errors and analyze its asymptotic properties. This estimator is required for the estimation of the variance function from noisy measurements and it works for sparse functional data ([11]) as well. Numerical performance reveals that the new estimator outperforms the benchmark method in [11].

References

[1] Dawson, M. and Müller, H.-G. (2018), 'Dynamic modeling of conditional quantile trajectories, with application to longitudinal snippet data', *Journal of the American Statistical Association* **113**(524), 1612–1624.

[2] Delaigle, A. and Hall, P. (2013), 'Classification using censored functional data', *Journal of the American Statistical Association* **108**(504), 1269–1283.

[3] Delaigle, A. and Hall, P. (2016), 'Approximating fragmented functional data by segments of Markov chains', *Biometrika* **103**(4), 779–799.

[4] Descary, M.-H. , and Panaretos, V. M. (2019), 'Recovering covariance from functional fragments', *Biometrika* **106**(1), 145–160.

[5] Kneip, A. and Liebl, D. (2018), 'On the optimal reconstruction of partially observed functional data', *arXiv*.

[6] Kraus, D. (2015), 'Components and completion of partially observed functional data', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77**(4), 777–801.

[7] Li, Y. and Hsing, T. (2010) , 'Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data', *The Annals of Statistics* **38**(6), 3321–3351.

[8] Liebl, D. and Rameseder, S. (2019), 'Partially observed functional data: The case of systematically missing parts', *Computational Statistics & Data Analysis* **131**, 104–115.

[9] Mojirsheibani, M. and Shaw, C. (2018), 'Classification with incomplete functional covariates', *Statistics & Probability Letters* **139**, 40–46.

[10] Stefanucci, M., Sangalli, L. M., and Brutti, P. (2018), 'PCA-based discrimination of partially observed functional data, with an application to aneurisk65 data set', *Statistica Neerlandica* **72**(3), 246–264.

[11] Yao, F., Müller, H.-G., and Wang, J.-L. (2005), 'Functional data analysis for sparse longitudinal data', *Journal of the American Statistical Association* **100**(470), 577–590.

[12] Zhang, A. and Chen, K. (2018) , 'Functional data approaches for mixed longitudinal studies, with applications in midlife women's health', *arXiv*.

[13] Zhang, X. and Wang, J. L. (2016) , 'From sparse to dense functional data and beyond', *The Annals of Statistics* **44**, 2281–2321.

Relative perturbation bounds with applications to empirical covariance operators

Moritz Jirak
(joint work with Martin Wahl)

The empirical covariance operator is a central object in high-dimensional probability. An important question studied in this context is the behaviour of empirical eigenvalues and eigenvectors. Knowing that the empirical covariance operator $\hat{\Sigma}$ is close to the population one Σ, one wants to infer that the empirical eigenvalues and eigenvectors do not deviate too much from their population counterparts. There is, by now, quite an extensive literature in this area regarding stochastic perturbation bounds. Roughly speaking, the assumptions of most results can be classified into a stochastic and an algebraic part.

- Stochastic: Given a sample $X_1, \ldots, X_n$, it is usually assumed that the sequence is i.i.d. Moreover, expressing $X_i = \sum_{j \geq 1} \sqrt{\lambda_j} u_j \eta_{ij}$ by its Karhunen-Loève expansion with eigenvalues $(\lambda_j)_{j\geq 1}$ and eigenvectors $(u_j)_{j \geq 1}$, an often made key assumption is that the coefficients $(\eta_{ij})_{j\geq 1}$ are independent and sub-Gaussian.
- Algebraic: The relation between the eigenvalues and their size is a key feature and immanent to the problem. It is typically expressed in terms of spectral gaps, growth or decay rates of the eigenvalues and sometimes linked to the dimension $d = \dim \mathcal{H}$ of the underlying Hilbert space $\mathcal{H}$ (or to some other notion of dimension of the underlying distribution) and the sample size n.

Our main objective is to circumvent most of these kind of conditions and develop relative perturbation results subject only to very little assumptions. As applications, we obtain concentration inequalities and central and non-central limit theorems. Our results apply to stationary sequences that may be weakly dependent or even exhibit long-memory, given very mild moment assumptions. Moreover, we allow for any kind of dependence relation between the coefficients $(\eta_{ij})_{j\geq 1}$, in particular, no independence is required. Regarding the underlying algebraic structure, we show that a basic quantity is given by the function

$$j \mapsto \mathbf{r}_j(\Sigma) = \sum_{k \neq j} \frac{\lambda_k}{|\lambda_j - \lambda_k|} + \frac{\lambda_j}{g_j}, \tag{1}$$

which we refer to as the *relative rank* of Σ (we actually consider a generalisation with multiplicities). In (1), g_j denotes the j-th spectral gap defined by $g_j = \min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})$ for $j \geq 2$ and $g_1 = \lambda_1 - \lambda_2$. It turns out that this function gives rise to necessary and sufficient conditions for some of our results.

The study of general perturbation bounds has a long tradition in matrix analysis, functional analysis, and operator theory. Classical perturbation bounds for eigenvalues and eigenspaces include the Weyl inequality and the Davis-Kahan $\sin \Theta$ theorem, see e.g. [2]. These bounds have been extended in many directions. A basic tool in perturbation theory for linear operators is the holomorphic functional calculus [6]. Key ingredients such as Cauchy's integral formula and the resolvent equations have been successfully applied to various stochastic perturbation problems, see e.g. [7, 3, 9] to mention a few. Typical (absolute) stochastic perturbation results for $\hat{\Sigma}$ state that empirical and population eigenvalues or eigenvectors are close to each other if some norm (usually the operator norm) of $\hat{\Sigma} - \Sigma$ is small.

Regarding random matrices, a fundamental question is to find precise estimates of corresponding norms. A number of more recent results established tight bounds for the operator norm of (possibly structured) random matrices, see for instance [1, 10]. However, all those and related results do not apply to empirical covariance operators, which, due to their quadratic structure, are fundamentally different objects. Using the method of generic chaining (cf. [11]), it has been recently shown in [8] that for sub-Gaussian i.i.d. observations the size of $\|\hat{\Sigma} - \Sigma\|_\infty$ is characterised by $\|\Sigma\|_\infty$ and the effective rank $\mathbf{r}(\Sigma) = \mathrm{tr}(\Sigma)/\|\Sigma\|_\infty$. Moving to a more special setup, a precise characterisation of the operator norm is possible in terms of the Tracy-Widom law, see for instance [4, 5].

The goal here is to develop tight relative perturbation bounds, by going significantly beyond the operator norm $\|\cdot\|_\infty$. This is achieved by exploiting a new contraction property for empirical spectral projectors. We require two ingredients. For the probabilistic part, we demand that certain relative coefficients of $\hat{\Sigma} - \Sigma$ are small. This allows us to avoid restrictive probabilistic assumptions like sub-Gaussianity and independence of the coefficients $(\eta_{ij})_{j\geq 1}$ in our applications. For the algebraic structure, we formulate conditions in terms of the relative rank. This allows us to circumvent absolute quantities related to the size of $\|\hat{\Sigma} - \Sigma\|_\infty$ like the effective rank. Although our approach is motivated from stochastic fluctuations and their properties, the results are equally valid for deterministic perturbations and are by no means restricted to empirical covariance operators.

References

[1] A.S. Bandeira and R. van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.*, 44(4):2479–2506, 2016.

[2] R. Bhatia. *Matrix analysis*. Springer-Verlag, New York, 1997.

[3] N. Hilgert, A. Mas, and N. Verzelen. Minimax adaptive tests for the functional linear model. *Ann. Statist.*, 41(2):838–869, 04 2013.

[4] I. M. Johnstone. On the distribution of the largest principal component. *Ann. Statist.*, 29:295–327, 2000.

[5] N. El Karoui. Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.*, 35(2):663–714, 03 2007.

[6] T. Kato. *Perturbation theory for linear operators*. Springer-Verlag, Berlin, reprint of the 1980 edition, 1995.

[7] V. Koltchinskii and E. Giné. Random matrix approximation of spectra of integral operators. *Bernoulli*, 6:113–167, 2000.

[8] V. Koltchinskii and K. Lounici. Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. Henri Poincaré*, 52:1976–2013, 2016.

[9] V. Koltchinskii and K. Lounici. Normal approximation and concentration of spectral projectors of sample covariance. *Ann. Statist*, 45:121–157, 2017.

[10] R. Latała, R. van Handel, and P. Youssef. The dimension-free structure of nonhomogeneous random matrices. *ArXiv e-prints*, 2017.

[11] M. Talagrand. *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics. Springer Berlin Heidelberg, 2016.

High-Dimensional Functional Factor Models

Shahin Tavakoli
(joint work with Marc Hallin, Gilles Nisol)

In this paper, we set up theoretical foundations for high-dimensional functional factor models for the analysis of large panels of functional time series (FTS). We first establish a representation result stating that if the first r eigenvalues of the covariance operator of the cross-section of N FTS are unbounded as N diverges and if the $(r+1)$-th eigenvalue is bounded, then we can represent each FTS as a sum of a common component driven by r factors, common to all the series, and a weakly cross-correlated idiosyncratic component (all the eigenvalues of the corresponding covariance operator bounded as $N \to \infty$). Our model and theory are developed in a general Hilbert space setting that allows panels mixing functional and scalar time series. We then turn to the estimation of the factors, their loadings, and the common components. We derive consistency results in the asymptotic regime where the number N of series and the number T of time observations diverge, thus exemplifying the "blessing of dimensionality" that explains the success of factor models in the context of high-dimensional (scalar) time series. Our results encompass the scalar factor models, for which they reproduce and extend, under weaker conditions, well-established results ([2, 3, 1, 5, 6]). We provide numerical illustrations that corroborate the convergence rates predicted by the theory, and provide finer understanding of the interplay between N and T for estimation purposes. We conclude with an empirical illustration on a dataset of intraday S&P100 and Eurostoxx 50 stock returns, along with their scalar overnight returns. A preprint of the paper is available on Arxiv ([4]).

References

[1] Bai, J. & Ng, S. (2002), *Determining the number of factors in approximate factor models*, Econometrica **70**(1), 191–221.

[2] Chamberlain, G. (1983), *Funds, factors, and diversification in arbitrage pricing models*, Econometrica **51**, 1281–1304.

[3] Chamberlain, G. & Rothschild, M. (1983), *Arbitrage, factor structure, and mean-variance analysis on large asset markets*, Econometrica **51**, 1305–1324.

[4] Nisol, G., Tavakoli, S. & Hallin, M. (2019), *High-dimensional functional factor models*, arXiv preprint arXiv:1905.10325.

[5] Stock, J. H. & Watson, M. W. (2002), *Forecasting using principal components from a large number of predictors*, J. Am. Stat. Assoc. **97**(460), 1167–1179.

[6] Stock, J. H. & Watson, M. W. (2002), *Macroeconomic forecasting using diffusion indexes*, J. Bus. Econ. Stat. **20**(2), 147–162.

Testing relevant hypotheses in functional data via self normalization

Holger Dette
(joint work with Kevin Kokot, Stanislav Volgushev)

Most of the available statistical methodology on testing statistical hypotheses in functional data considers hypotheses of the form

$$H_0 : d = 0 \text{ versus } H_1 : d \neq 0 \tag{1}$$

where d is a real valued parameter such as the norm of the mean function in one sample or the norm of the difference of two mean functions or two covariance operators from two samples. For independent data the quantiles for corresponding tests can be easily obtained by asymptotic theory as the unknown quantities in the limit distribution of the test statistics can be reliably estimated. However, for functional samples exhibiting temporal dependence, the asymptotic distribution of many commonly used tests statistics involves the long-run variance, which makes the statistical inference substantially more difficult as good estimates of the long-run variance are required. As alternative (asymptotically) pivotal test statistics based on the concept of self-normalization can be obtained and these methods have recently been developed for the specific needs of functional data by [3] and [4] [see also [2] for a recent review].

A common feature of of most of the literature is that they usually address hypotheses of the form (1). However, in many applications one might not be interested in detecting very small deviations of the parameter d from 0 (often the researcher even knows that d is not exactly equal to 0, before any experiments have been carried out). Therefore we argue that one should carefully think about the size of the difference in which one is interested. In particular we propose to replace the hypotheses (1) by the hypotheses of a *relevant difference*, that is

$$H_0 : d \leq \Delta \quad \text{versus} \quad H_1 : d > \Delta\,, \tag{2}$$

where Δ is a pre-specified constant representing the "maximal" value for the parameter d, which can be accepted as not scientifically significant.

In this paper we discuss the problem of testing relevant hypotheses in the context of functional dependent data. We are particularly interested in methods based on self-normalization in order to avoid estimation of the long-run variance or resampling methods. For this purpose we modify the classical approaches to self-normalization based testing. This modification is of independent interest besides the field of functional data analysis and applicable in many other problems.

To be more precise let $L^2([0,1])$ denote the Hilbert space of square integrable functions on the set $[0,1]$ with the usual inner product $\langle\cdot,\cdot\rangle$ and corresponding norm $\|\cdot\|$. Let $\{X_n\}_{n\in\mathbb{Z}}$ denote a strictly stationary functional time series where the random variables X_n are elements in $L^2([0,1])$ with expectation $\mu := \mathbb{E}[X_1] \in L^2([0,1]$. Based on a sample $X_1,...,X_n$ we are interested in relevant hypotheses regarding the parameter $d = \|\mu\|^2 = \int_{[0,1]} \mu^2(t)dt$, that is

$$H_0 : \|\mu\|^2 \leq \Delta \quad \text{versus} \quad H_1 : \|\mu\|^2 > \Delta\,. \tag{3}$$

Define the partial sums

$$S_n(t,\lambda) := \frac{1}{n}\sum_{j=1}^{\lfloor n\lambda\rfloor} X_j(t)\,, \quad \lambda \in [0,1]\,,$$

then, under suitable assumptions, the statistic $\|S_n(\cdot,1)\|^2$ is a consistent estimator of $\|\mu\|^2$. Consequently a test for the hypotheses (3) is obtained by rejecting the null hypothesis for large values of

$$\hat{\mathbb{T}}_n = \|S_n(\cdot,1)\|^2 = \int_{[0,1]} S_n^2(t,1)dt\,. \tag{4}$$

Let ν denote a measure on the interval $[0,1]$ and define

$$\hat{\mathbb{V}}_n := \Big(\int_0^1 \Big[\int_{[0,1]} S_n^2(t,\lambda)dt - \lambda^2\int_{[0,1]} S_n^2(t,1)dt\Big]^2 \nu(d\lambda)\Big)^{1/2}\,. \tag{5}$$

We prove that under suitable assumptions

$$\Big(\sqrt{n}(\hat{\mathbb{T}}_n - d), \sqrt{n}\,\hat{\mathbb{V}}_n\Big) \overset{\mathcal{D}}{\to} \Big(\tau\mathbb{B}(1), \tau\Big(\int_0^1 \lambda^2(\mathbb{B}(\lambda)-\lambda\mathbb{B}(1))^2\nu(d\lambda)\Big)^{1/2}\Big)\,, \tag{6}$$

where $\mathbb{B}$ denotes a standard Brownian motion on the interval $[0,1]$ and τ is a non-negative constant. If $q_{1-\alpha}(\mathbb{W})$ denotes the $1-\alpha$ quantile of the distribution of the pivotal random variable

$$\mathbb{W} := \frac{\mathbb{B}(1)}{\big(\int_0^1 \lambda^2(\mathbb{B}(\lambda)-\lambda\mathbb{B}(1))^2\nu(d\lambda)\big)^{1/2}} \tag{7}$$

then we prove that the test which rejects the null hypothesis in (3), whenever

$$\hat{\mathbb{T}}_n > \Delta + q_{1-\alpha}(\mathbb{W})\hat{\mathbb{V}}_n\,, \tag{8}$$

is a consistent and asymptotic level α test. Details and proofs of these results can be found in [1].

References

[1] Dette, H., K. Kokot, and S. Volgushev (2018). Testing relevant hypotheses in functional time series via self-normalization. *arXiv:1809.06092*.

[2] Shao, X. (2015). Self-normalization for time series: A review of recent developments. *Journal of the American Statistical Association*, 110(512):1797–1817.

[3] Zhang, X., Shao, X., Hayhoe, K., and Wuebbles, D. J. (2011). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electron. J. Statist.*, 5:1765–1796.

[4] Zhang, X. and Shao, X. (2015). Two sample inference for the second-order property of temporally dependent functional data. *Bernoulli*, 21(2):909–929.

Object oriented data analysis of samples of networks

Ian Dryden
(joint work with Simon P. Preston, Katie E. Severn)

The topic of Object Oriented Data Analysis (OODA) began with [9], and a broad overview of the field with many examples has been given by [6]. Important aspects of OODA include the need to make choices about *i*) what the data objects are, *ii*) the conceptual space in which the data objects lie, and *iii*) the feature space that is used for practical data analysis.

Covariance functions and networks are types of object data that are used in many applications including in the analysis of spoken and written language. It is of interest to develop statistical techniques to compare samples of such data objects. In both cases one can choose the features for statistical analysis to be high-dimensional symmetric positive semi-definite matrices, where networks are represented by graph Laplacians, which is a subspace of the space of covariance matrices.

For the comparison of certain types of object data [1] and [7] considered the family of power Euclidean distances between pairs of covariance matrices and infinite dimensional covariance operators, respectively. In particular the power Euclidean metric between covariance matrices A and B is

$$d_\alpha(A, B) = \|A^\alpha - B^\alpha\|, \tag{1}$$

where $\|A\|$ is the Frobenius norm of A, and $A^\alpha = U\Lambda^\alpha U^T$ is the symmetric matrix power where $A = U\Lambda U^T$ is the usual spectral decomposition. A common choice is $\alpha = \frac{1}{2}$, which gives the symmetric matrix square root.

In text-based corpus analysis, word collocations are widely studied ([3]), i.e., words that have a tendency to co-occur; and text documents represented as word-pair co-occurrence counts can be identified as networks ([8]). Analysis of networks is a type of OODA analysis, with wide applications in neuroscience and genetics, besides text analysis. Let $G_m = (V, E)$, comprise a set of nodes, $V = \{v_1, v_2, \ldots, v_m\}$, and a set of edge weights, $E = \{w_{ij} : w_{ij} \geq 0, 1 \leq i, j \leq m\}$, indicating nodes v_i and v_j are either connected by an edge of weight $w_{ij} > 0$, or else unconnected if $w_{ij} = 0$, and suppose $w_{ij} = w_{ji}$ and $w_{ii} = 0$ (network is undirected and without loops). Any such network can be identified with its $m \times m$ graph Laplacian matrix $\mathbf{L} = (l_{ij})$, defined as

$$l_{ij} = \begin{cases} -w_{ij}, & \text{if } i \neq j \\ \sum_{k \neq i} w_{ik}, & \text{if } i = j \end{cases}$$

for $1 \leq i, j \leq m$. The space of graph Laplacians is a subset of the cone of symmetric positive semi-definite matrices ([4]).

[8] described a framework for manifold value data analysis of networks that uses the Euclidean power distance (1) and introduce a unique projection from the space of covariance matrices to the subspace of graph Laplacians. The projection can be computed efficiently using quadratic programming. [8] define embeddings, tangent

spaces and use the metrics and projection to perform extrinsic statistical analysis, such as calculating a mean of a sample of networks. The framework has similarities to the use of extrinsic methods in statistical shape analysis ([2]). Further statistical analysis such as regression and principal components analysis has been developed, and a hypothesis test has been described for testing the equality of means between two samples of networks.

In this presentation the methodology is described and applied to the set of novels by Jane Austen and Charles Dickens from the University of Birmingham CLiC project ([5]), illuminating striking differences in the way the novelists used words, and how their word usage changed over time.

References

[1] Dryden, I. L., Koloydenko, A., and Zhou, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.*, 3(3):1102–1123.

[2] Dryden, I. L. and Mardia, K. V. (2016). *Statistical Shape Analysis, with Applications in R, 2nd edition.* Wiley, Chichester.

[3] Gablasova, D., Brezina, V., and McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67:155–179.

[4] Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S., and Kolaczyk, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Stat.*, 11(2):725–750.

[5] Mahlberg, M., Stockwell, P., de Joode, J., Smith, C., and O'Donnell, M. B. (2016). CLiC Dickens: novel uses of concordances for the integration of corpus stylistics and cognitive poetics. *Corpora*, 11(3):433–463.

[6] Marron, J. S. and Alonso, A. M. (2014). Overview of object oriented data analysis. *Biometrical Journal*, 56(5):732–753.

[7] Pigoli, D., Aston, J. A. D., Dryden, I. L., and Secchi, P. (2014). Distances and inference for covariance operators. *Biometrika*, 101(2):409–422.

[8] Severn, K. E., Dryden, I. L., and Preston, S. P. (2019). Manifold valued data analysis of samples of networks, with applications in corpus linguistics. arXiv:1902.08290.

[9] Wang, H. and Marron, J. S. (2007). Object oriented data analysis: sets of trees. *Ann. Statist.*, 35(5):1849–1873.

The Procrustes metric on covariance operators is optimal transport: Statistical implications

Yoav Zemel

(joint work with Valentina Masarotto, Victor M. Panaretos)

Covariance operators are fundamental in functional data analysis, providing the canonical means to analyse functional variation via the celebrated Karhunen–Loève expansion. These operators may themselves be subject to variation, for instance in contexts where multiple functional populations are to be compared. Statistical techniques to analyse such variation are intimately linked with the choice of metric on covariance operators, and the intrinsic infinite-dimensionality of these operators.

Early attempts to tackle this problems utilise the fact that any covariance operator $S : \mathcal{H}$ to $\mathcal{H}$ (defined on a separable Hilbert space $\mathcal{H}$) is Hilbert–Schmidt, allowing to embed the space of covariance operators in the Hilbert space of Hilbert–Schmidt operators (see e.g., [6, 1]). This choice of distance, however, is effectively an extrinsic distance that does not take into account the nonnegative nature of covariance operators.

[9] were the first to consider alternative distances that are adapted to the geometry of the space of covariance operators. They consider the so-called Procrustes distance

$$\Pi(S_1, S_2) = \inf_{U^*U=\mathcal{I}} \left|\left|\left| S_1^{1/2} - S_2^{1/2} U \right|\right|\right|_2,$$

where $\mathcal{I}$ is the identity operator, $*$ is the adjoint, and $|||\cdot|||_2$ is the Hilbert–Schmidt norm, generalising the matrix case considered in [3]. For brevity we refer to the resulting metric space as the *Procrustes space.*

In our work [4, 5] we identify the Procrustes metric Π with the 2-Wasserstein metric (see [7, 8] for a recent review and a book, aimed at a statistical audience) between the centred Gaussian measures with covariances S_1 and S_2 and exploit this in order to construct a powerful test of homogeneity and for analysing variation using principal component analysis in the Procrustes space. A key component is the linear operator, *transport map*,

$$\mathbf{t}_{S_1}^{S_2} := S_1^{-1/2}[S_1^{1/2} S_2 S_1^{1/2}]^{1/2} S_1^{-1/2}.$$

It exists when $\ker(S_1) \subseteq \ker(S_2)$, in which case it is self-adjoint, nonnegative and possibly unbounded. This map can be thought of as "deforming" a template process $X_1 \sim N(0, S_1)$ to a "warped" process $X_2 \sim N(0, S_2)$ because $\mathbf{t}_{S_1}^{S_2} X_1$ has the same law as X_2. The relation to the Procrustes distance is

$$\mathbb{E} \left\| \mathbf{t}_{S_1}^{S_2} X_1 - X_1 \right\|^2 = \Pi^2(S_1, S_2).$$

The warping nature of the transport maps is associated with a data-generating mechanism in Procrustes space via phase variation. Let S be a template covariance and $\mathbf{t} : \mathcal{H} \to \mathcal{H}$ a random self-adjoint nonnegative operator with mean identity. Then under suitable conditions, S is a Fréchet mean of its perturbed version $\mathbf{t}S\mathbf{t}$ with respect to Procrustes distance: it minimises the Fréchet functional

$$R \mapsto F(R) = \mathbb{E}[\Pi^2(R, \mathbf{t}S\mathbf{t})]. \tag{1}$$

Note that if $X \sim N(0, S)$, then $\mathbf{t}X \sim N(0, \mathbf{t}S\mathbf{t})$, so the random $\mathbf{t}S\mathbf{t}$ expresses that $\mathbf{t}$ acts on the underlying Hilbert space $\mathcal{H}$.

Topology. Convergence of S_n to S with respect to the Procrustes distance is equivalent to

- Convergence in distribution of the Gaussian measures $N(0, S_n)$ to $N(0, S)$;
- Convergence with respect to trace norm;
- Convergence of $S_n^{1/2}$ to $S^{1/2}$ in Hilbert–Schmidt norm.

When S_n are finite-dimensional projections of S with respect to a given basis, the rate of convergence can be quantified, and it is uniform when S ranges over suitable classes of covariance operators.

Geometry. The tangent space at S is the Hilbert space

$$\mathrm{Tan}_S = \overline{\{A : A = A^*, \left|\left|\left| S^{1/2}A \right|\right|\right|_2 < \infty\}},$$

where the closure is with respect to the associated inner product

$$\langle A, B\rangle_S = \mathrm{tr}[ASB] = \mathbb{E}\langle AX, BX\rangle, \qquad X \sim N(0, S).$$

In view of the compactness of $S^{1/2}$, this space includes all bounded self-adjoint operators A, but also certain unbounded ones. For example, if $S^{1/2}$ is trace-class, then $S^{1/4}$ is Hilbert–Schmidt and the unbounded operator $A = S^{-1/4}$ is also in the tangent space.

The exponential map at S (from Tan_S to the Procrustes space) is

$$\exp_S[A] = (A + \mathcal{I})S(A + \mathcal{I}).$$

When $\mathbf{t}_{S_0}^{S_1}$ exists, its difference from the identity is the log map

$$\log_{S_0}(S_1) = \mathbf{t}_{S_0}^{S_1} - \mathcal{I},$$

and there is a unique minimal constant speed geodesic given by

$$S_s = s^2 S_1 + (1-s)^2 S_0 + s(1-s)[\mathbf{t}_0^1 S_0 + S_0 \mathbf{t}_0^1], \qquad s \in [0, 1].$$

Fréchet means. Fréchet means of a collection of covariances $S_1, \ldots, S_n$ are defined in analogy with (1). One can show that a Fréchet mean $\overline{S}$ always exists, and uniquely so if at least one covariance S_i is injective. The mean is also stable, in the sense that if $\Pi(S_i^k, S_i) \to 0$ for all i as $k \to \infty$, then Fréchet means of $(S_1^k, \ldots, S_n^k)$ converges to $\overline{S}$. This covers, in particular, the most important case where S_i^k are projections of S_i to k-dimensional subspaces that approximate $\mathcal{H}$. Although it has no closed-form formula in general, $\overline{S}$ can be computed efficiently using a Procrustes-type steepest descent algorithm [4, Section 8].

Test of homogeneity. We propose a new test statistic for the null hypothesis

$$H_0 : \quad S_1 = \cdots = S_n$$

on the basis of samples of curves $X_{i,1}, \ldots, X_{i,n_i}$ from processes with covariance operators S_i. The key step is to recast the H_0 as

$$H_0 : \quad \mathbf{t}_{\overline{S}}^{S_1} = \cdots = \mathbf{t}_{\overline{S}}^{S_n} = \mathcal{I},$$

where $\overline{S}$ is a Fréchet mean of $S_1, \ldots, S_n$. One then estimates each S_i by the empirical covariances $\widehat{S_i}$ from which $\widehat{\overline{S}}$ and the transport maps $\widehat{\mathbf{t}}_i = \mathbf{t}_{\widehat{\overline{S}}}^{\widehat{S_i}}$ are constructed. We show that these transport maps are not only well-defined, but also exist as *bounded* linear operators. The test statistic is (weighted versions can readily be considered)

$$T = \sum_{i=1}^{n} \left|\left|\left| \widehat{\mathbf{t}}_i - \mathcal{I} \right|\right|\right|,$$

where the norm could be e.g., the operator norm, Hilbert–Schmidt norm or trace norm. The null is rejected for large values of T, calibrated by permutations. We observe in a myriad of different scenarios that the test based on T overpowers the state of the art procedure of [2], which uses directly the distance Π.

Functional covariance analysis of variance. If H_0 is rejected, one could aim to describe and interpret the main modes of variation in the collection of covariances $S_1, \dots, S_n$. This can be achieved by carrying out the principal component analysis at the level of the tangent space. One considers the collection $\mathbf{t}_{S}^{S_i}$, $i = 1, \dots, n$ as elements in the Hilbert space of operators with the inner product $\langle \cdot, \cdot \rangle_S$. Some care needs to be taken in the computations, since this norm is not the standard Hilbert–Schmidt norm.

References

[1] Boente, G., Rodriguez, D., Sued, M., *Testing equality between several populations covariance operators*, Annals of the Institute of Statistical Mathematics **70**(2018), 919–950.

[2] Cabassi, A., Pigoli, D., Secchi, P., Carter, P. A., *Permutation tests for the equality of covariance operators of functional data with applications to evolutionary biology*, Electronic Journal of Statistics **11** (2017), 3815–3840.

[3] Dryden, I. L., Koloydenko., A. Zhou, D., *Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging*, Annals of Applied Statistics **3** (2009), 1102–1123.

[4] Masarotto, V., Panaretos V.M., Zemel, Y., *Procrustes Metrics on Covariance Operators and Optimal Transportation of Gaussian Processes*, Invited Paper, Special Issue on Manifold Statistics, Sankhyā A (in press).

[5] Masarotto, V., Panaretos, V.M., Zemel, Y., *Transportation-Based Functional ANOVA and PCA for Covariance Operators*, in preparation.

[6] Panaretos, V. M., Kraus, D., Maddocks, J. H., *Second-order comparison of Gaussian random functions and the geometry of DNA minicircles*, Journal of the American Statistical Association **105** (2010), 670–682.

[7] Panaretos, V. M., Zemel, Y., *Statistical Aspects of Wasserstein Distances*, Annual Review of Statistics and Its Applications **6** (2019), 405–431.

[8] Panaretos, V. M., Zemel, Y., An Invitation to Statistics in Wasserstein Space, *SpringerBriefs in Probability & Mathematical Statistics*, forthcoming.

[9] Pigoli, D., Aston, J. A., Dryden, I. L., Secchi, P., *Distances and inference for covariance operators*, Biometrika **101** (2014), 409–422.

Data Integration Via Analysis of Subspaces (DIVAS)

J.S. Marron

A major challenge in the age of Big Data is the integration of disparate data types into a data analysis. That is tackled here in the context of data blocks measured on a common set of experimental cases. This data structure motivates the simultaneous exploration of the joint and individual variation within each data block. DIVAS improves earlier methods using a novel random direction approach to statistical inference, and by treating partially shared blocks. Usefulness is illustrated using mortality, cancer and neuroimaging data sets. This improves upon the earlier JIVE methodology of Lock et al. [2] and the AJIVE proposed by Feng et al. [1]. It lies in the general area of Object Oriented Data Analysis as defined by Wang and Marron [4], and more recently overviewed in Marron and Alonso [3].

References

[1] Qing Feng, Meilei Jiang, Jan Hannig, and J. S. Marron. Angle-based joint and individual variation explained. *J. Multivariate Anal.* **166**:241–265, 2018.

[2] Eric F. Lock, Katherine A. Hoadley, J. S. Marron, and Andrew B. Nobel. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *Ann. Appl. Stat.* **7**(1):523–542, 2013.

[3] J. S. Marron and Andrés M. Alonso. Overview of object oriented data analysis. *Biom. J.* **56**(5):732–753, 2014.

[4] Haonan Wang and J. S. Marron. Object oriented data analysis: sets of trees. *Ann. Statist.* **35**(5):1849–1873, 2007.

Simplicial feature maps and Random Euler measures

Katharine Turner
(joint work with Kathryn Hess, Victor Panaretos, Gard Spreemann)

One of the great challenges with statistically analysing complex data (such as simplicial complexes) is the need to first map them to a common space. Often this is done via computing various summary statistics and then statistically analysing these summary statistics instead of the raw complex object.

The Euler characteristic is a classical topological invariant that bridges many different areas of topology and geometry. The Euler characteristic, denoted χ, is a topological invariant that appears in many different areas of mathematics. It was originally defined for polyhedra according to the formula $\chi = V - E + F$ where V is the number of vertices (0-cells), E the number of edges (1-cells) and F the number of faces (2-cells). There are now many equivalent ways of computing the Euler characteristic in different setting. This formula extends easily to all simplicial complexes as the alternative sum of the number of cells in each dimension;

$$\chi(K) = \sum_{k=0}^{\dim K} (-1)^k (\text{ number of } k \text{ - cells.})$$

A key property is that the Euler characteristic is independent of triangulations; that is if we consider the same subset of space as two different geometric simplicial complexes then their Euler characteristics will agree.

The main disadvantage of the Euler characteristic for data analysis is that, for a given simplical complex, the Euler characteristic is only a single number and we would like to compute more information. To this end, we will define a generalisation of the Euler characteristic which will also use location information. For a region we can consider the restriction of the simplical complex to that region and compute the Euler characteristic of this restriction. Appropriately defined, this process will determine a finitely additive signed measure. A finitely additive signed measure over an algebra Σ is analogously a function

$$\mu : \Sigma \to \mathbb{R} \cup \{\infty, -\infty\}$$

such that $\mu(\emptyset) = 0$ and μ is finitely additive, that is, it satisfies the equality

$$\mu\left(\bigcup_{n=1}^{k} A_n\right) = \sum_{n=1}^{k} \mu(A_k)$$

for any finite sequence $A_1, A_2, \ldots, A_k$ of disjoint sets in Σ. If Σ is a finite σ-algebra then finite additivity implies countable additivity and thus any finitely-additive measure μ over Σ is also a measure.

Given a simplicial complex and a feature map given over its vertex set into an affine feature space, we can construct a map over the entire simplicial complex using linear interpolation. We call this a simplicial feature map. We can then construct a finitely additive signed measure over the feature space by considering Euler characteristics (with compact supports) of the preimage of the simplicial feature map.

We define a topological summaries in the form of an Euler measure over a *common* relevant algebra of subsets of the feature space. This space of additive measures over this algebra is a vector space. One basis is the measures over the minimal sets of the algebra. We can consider each topological summary statistics as a vector. Given multiple instances we can perform statistical and machine learning procedures to analyse the sets of vectors. Examples include regression and covariance analysis.

The motivation is to develop methods for understanding simplical complexes where the vertices have location or other relavent information. In particular we want to analyse the Blue Brain microcircuit. From this microcircuit we can build a simplicial complex where each vertex represents an individual neuron, the edges are synaptic connections, and higher dimensional simplices correspond to cliques of information flow. There are two different types of spacial information that we will study. The first is where the physical location of the neuron, and the second is the location of the neuron in a feature space of firing patterns when the microcircuit in simulations.

In an example application we perform a preliminary analysis of data from the Blue Brain Project. We investigate the relationship between the Euler measures of different regions, giving insight into the structure of the brain. We also can compare the Euler measures of the different microcircuits for different groups of input and for different stimuli of the microcircuit.

On central limit theorems on manifolds and stratified spaces

Stephan Huckemann
(joint work with Benjamin Eltzner)

1. From the BP-CLT to the EH-CLT

1.1. **General Setup.** Consider random elements $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} X$ on a topological space Q called the *data space* that is linked to a topological space P called the *descriptor space* via a continuous map $\rho : Q \times P \to [0, \infty)$. This gives rise to the *empirical* and *population Fréchet function*

$$F_n(p) \;=\; \frac{1}{n}\sum_{j=1}^{n} \rho(X_j, p)\,, \quad F(p) \;=\; \mathbb{E}\left[\rho(X, p)\right]\,,$$

respectively. The set of their minimizers

$$E_n \;=\; \operatorname*{argmin}_{p\in P} F_n(p)\,, \quad E \;=\; \operatorname*{argmin}_{p\in P} F(p)\,,$$

are called *sample* and *population Fréchet means*, respectively.

Together with a *loss function* $d : P \times P \to [0, \infty)$, i.e. d is continuous and $d(p, p') = 0$ if and only if $p = p'$, consider the condition

$$(A1) \quad \begin{array}{l} \text{there is a constant } C > 0 \text{ such that } |\rho(q, p) - \rho(q, p')| \leq C\, d(p, p') \\ \text{for all } q \in Q,\ p, p' \in P\,. \end{array}$$

Note that in case of $d = \sqrt{\rho}$ being a distance on $Q = P$, due to the triangle inequality, (A1) is valid with $C = 1$.

Theorem 1 ([7]). *With the above setup, if P is separable, $F(p) < \infty$ for all $p \in P$ and if (A1) holds, then $\bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} E_k} \subseteq E$ a.s.* .

Remark 2. *For a quasi-metric $d = \sqrt{\rho}$ on $Q = P$, this has been proven by* [15]. *If $E \neq \emptyset$, if $\overline{\bigcup_{k=n}^{\infty} E_k}$ is d-Heine-Borel (i.e. every closed and d-bounded subset is compact) and if (ρ, d) is coercive (as detailed in* [7]*) then for every $\epsilon > 0$ there is a random $N(\epsilon) \in \mathbb{N}$ such that $E_n \subseteq \{p \in P : d(p, E) \leq \epsilon\}$ almost surely for all $n \geq N(\epsilon)$, cf.* [2, 7].

1.2. **Manifold Interlude.**

Theorem 3 ([3, 1]). *If $P = Q$ is a manifold with geodesic distance $d = \sqrt{\rho}$, $\{\mu\} = E$, $\mu_n \in E_n$ a measurable selection with $\mu_n \overset{a.s.}{\to} \mu$ and if $\phi : U \to V$ is a local chart near $\mu \in U \subset P$, $V \in \mathbb{R}^m$ for some $m \in \mathbb{N}$ with $\phi(\mu) = 0$ and $\phi(\mu_n) = x_n$, then, with the assumptions below,*

$$\sqrt{n}\; \phi^{-1}(\mu_n) \;\overset{\mathcal{D}}{\to}\; \mathcal{N}(0, H^{-1}\Sigma H^{-1})\,,$$

(A2): $x \mapsto \rho\big(X, \phi^{-1}(x)\big) \in \mathcal{C}^2(V)$ *a.s.,*
(A3): $\exists \operatorname{cov}\big[\operatorname{grad}|_{x=0}\rho\big(X, \phi^{-1}(x)\big)\big] = \Sigma,$

(A4): $\exists$ $\mathrm{Hess}|_{x=0} F\left(\phi^{-1}(x)\right) = H$ *and* $H > 0$,
(A5): $\mathbb{E}\left[\sup_{\|x\|\le\epsilon} \left|\mathrm{Hess}|_x \rho\left(\phi^{-1}(x)\right) - \mathrm{Hess}|_{x=0}\rho\left(\phi^{-1}(x)\right)\right|\right] \to 0$ *($\epsilon \to 0$).*

Proof. By definition with some random $\tilde{x}_n$ between x_n and 0, a.s.,

$$\begin{aligned} 0 &= \sqrt{n}\,\mathrm{grad}|_{x=x_n} F_n\left(\phi^{-1}(x)\right) \\ &= \underbrace{\sqrt{n}\,\mathrm{grad}|_{x=0} F_n\left(\phi^{-1}(x)\right)}_{\xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma)} + \underbrace{\mathrm{Hess}|_{x=\tilde{x}_n} F_n\left(\phi^{-1}(x)\right)}_{\xrightarrow{\mathbb{P}} H} \sqrt{n} x_n \end{aligned}$$

where the convergence of the first term is due to the classical central limit theorem, the convergence of the second term follows from the following argument. For every $\delta > 0$, by Chebyshev's inequality,

$$\begin{aligned} &\mathbb{P}\left\{\left|\mathrm{Hess}|_{x=\tilde{x}_n} F_n\left(\phi^{-1}(x)\right) - \mathrm{Hess}|_{x=0} F_n\left(\phi^{-1}(x)\right)\right| > \delta\right\} \\ \le\ & \frac{1}{\delta}\,\mathbb{E}\left[\left|\mathrm{Hess}|_{x=\tilde{x}_n} F_n\left(\phi^{-1}(x)\right) - \mathrm{Hess}|_{x=0} F_n\left(\phi^{-1}(x)\right)\right|\right] \\ \le\ & \frac{1}{\delta}\,\mathbb{E}\left[\left|\mathrm{Hess}|_{x=\tilde{x}_n} \rho\left(\phi^{-1}(x)\right) - \mathrm{Hess}|_{x=0} \rho\left(\phi^{-1}(x)\right)\right|\right] \\ \le\ & \frac{1}{\delta}\,\mathbb{E}\left[\sup_{\|x\|\le\|x_n\|}\left|\mathrm{Hess}|_{x} \rho\left(\phi^{-1}(x)\right) - \mathrm{Hess}|_{x=0} \rho\left(\phi^{-1}(x)\right)\right|\right] \ \to\ 0 \end{aligned}$$

due to (A5) because $\|x_n\| \overset{\text{a.s.}}{\to} 0$ by hypothesis. Now application of the classical strong law $\mathrm{Hess}|_{x=0} F_n\left(\phi^{-1}(x)\right) \overset{\text{a.s.}}{\to} \mathrm{Hess}|_{x=0} F\left(\phi^{-1}(x)\right)$ yields the assertion. □

Example 4 ([6]). *Consider $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X$ on $Q = \mathbb{S}^1 = [-\pi, \pi)$ where π is identified with $-\pi$ with unique mean $\mu = 0$ and sample mean $\mu_n = x_n$. Then, for $x > 0$ we have*

$$\begin{aligned} nF_n(x) &= \sum_{x-\pi\le X_j} (X_j - x)^2 + \sum_{x-\pi > X_j} (X_j + 2\pi - x)^2 \\ &= \sum_{j=1}^{n} (X_j - x)^2 + 4\pi \sum_{x-\pi>X_j} (X_j - x + \pi)^2 . \end{aligned}$$

If X features a density f near $\pm\pi$ with respect to the uniform measure, then (A2) is valid and $\mathrm{Hess}|_x F_n(x) = 2$ a.s. for x sufficiently small, but we have for n sufficiently large

$$0 = 2\sqrt{n}(x_n - \bar{X}) - \sqrt{n}\,\underbrace{\frac{4\pi}{n}\sum_{x_n-\pi>X_j} 1}_{\approx f(-\pi)\,x_n},$$

$$\text{hence, } \sqrt{n}x_n \underbrace{2\big(1 - 2\pi f(-\pi)\big)}_{\neq 2} = 2\sqrt{n}\bar{Y} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathrm{cov}[2\bar{Y}])$$

for $f(-\pi) > 0$, which is possible. So (A5) is not valid. It is even possible that $f(-\pi) = \frac{1}{2\pi}$, so that $H = 0$ and (A4) is no longer valid.

1.3. **Empirical Process Theory for the General Setup.** We assume $E = \{\mu\}$, $E_n \ni \mu_n$ measurable $\overset{\text{a.s.}}{\to}$ μ, P has a local manifold structure near μ with local chart ϕ, $\phi(\mu) = 0$, $2 \leq r \in \mathbb{N}$ and that

$$\begin{array}{lrcl}
(A1') & \big|\rho(X, p - \rho(X, p')\big| & \leq & \dot\rho(X)\, d(p,p'),\ \forall p, p' \in P \text{ near } \mu, \\
 & & & \text{with } \mathbb{E}\big[\dot\rho^2(X)\big] < \infty\,, \\
(A2') & \exists\, \text{grad}\,|_{x=0}\rho\big(X, \phi^{-1}(x)\big) & =: & \dot\rho_0(X) \ \text{ a.s. with existing } \text{cov}[\dot\rho_0(X)]\,, \\
(A4') & F\big(\phi^{-1}(x)\big) & = & F(\mu) + \sum_{k=1}^m \underbrace{T_k}_{>0} \big((Rx)_k\big)^r + o\big(\|x\|^r\big)\,.
\end{array}$$

Here R is a rotational matrix and $(Rx)_k$ denotes the k-th component. As in [14], one can show that there is a constant $C > 0$ such that

$$\begin{array}{rcl}
\sup\limits_{\|x\|<\delta} \big|F\big(\phi^{-1}(x)\big) - F(\mu)\big| & \leq & C\delta^\alpha \ (\alpha = r)\,, \\
\mathbb{E}\left[\sqrt{n} \sup\limits_{\|x\|<\delta} \big|F_n\big(\phi^{-1}(x)\big) - F\big(\phi^{-1}(x)\big) - (F_n(\mu) - F(\mu))\big|\right] & \leq & C\delta^\beta \ (\beta = 1)\,,
\end{array}$$

and that $n^{\frac{1}{2(\alpha-\beta)}} x_n = n^{\frac{1}{2(r-1)}} x_n = O_p(1)$. In consequence, one can show the following.

Theorem 5 ([5]). *Under the above assumptions with $T = \text{diag}(T_1, \dots, T_m)$,*

$$\sqrt{n}\,\underbrace{\text{sign}(Rx_n)|Rx_n|^{r-1}}_{\textit{componentwise}} \overset{\mathcal{D}}{\to} \mathcal{N}\left(0, T^{-1}\text{cov}[\dot\rho_0(X)]T^{-1}\right)\,.$$

Definition 6. *We say that the mean μ of a random variable as in Theorem 5 is $r - 2$ smeary.*

For manifolds $Q = P$ and $d = \sqrt{\rho}$ the intrinsic geodesic distance, it was shown in [6, 5] that

a) $\exists$ arbitrary smeariness on $\mathbb{S}^1$ and products thereof;
b) $\exists$ $r - 2 = 2$ smeariness on $\mathbb{S}^m$ for all $m \in \mathbb{N}$;
c) smeariness, although only present for a null set of the parameter space of distributions, influences finite sample rates nearby. This phenomenon is called *finite sample smeariness.*

2. Applications

Typical applications that have been developed to date (cf. [8, 7]) are intrinsic MANOVA and one- or two-sample tests for

(1) first geodesic principal component on manifolds and shape spaces,
(2) great/small subspheres in $\mathbb{S}^{m-1}$,
(3) and classical PCA.

Example 7. *We illustrate (3) by considering a random variable X on $Q = \mathbb{R}^m$ with $\mathbb{E}[X] = 0$ and existing* $\operatorname{cov}[X]$ *that allows the spectral decomposition* $\operatorname{cov}[X] = V\Lambda V^T$, $V \in O(m)$, $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_m)$ *with* $\lambda_1 = \dots = \lambda_k > \lambda_{k+1} \geq \dots \geq \lambda_m > 0$*. As an example, we treat here the asymptotics of the first k-dimensional eigenspace from classical PCA. Then the descriptor space is the Grassmannian* $P = G(m,k) \ni p = \operatorname{span}\underbrace{(v_{k+1}, \dots, v_m)}_{=:W}{}^{\perp}$.

With the analog spectral decomposition of the sample covariance $\operatorname{cov}[X_1, \dots, X_n] = \hat{V}\hat{\Lambda}\hat{V}^T$, $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_k \geq \hat{\lambda}_{k+1} \geq \dots \hat{\lambda}_m \geq 0$*, the distance* $d(p,p') = \min_{R \in O(m-k)} \|W - RW'\|$ *and link function* $\rho(X,p) = 1 - \|\sum_{j=k+1}^m v_k v_k^T X\|^2 = 1 - \|WW^TX\|^2 = 1 - \operatorname{tr}(W^TXX^TW)$*, we have*

$$\rho(X,p') - \rho(X,p) = \operatorname{tr}(W^TXX^TW) - \operatorname{tr}(W'^TXX^TW')$$

In consequence, we have (A1) and hence the strong law from Theorem 1. Further, we have (A4') with $r = 2$. In case of $E[\|X\|^4] < \infty$ *we have also (A1') and (A2'), so that Theorem 5 yields a $\sqrt{n}$-Gaussian CLT.*

Notably, there exists a version of a nested CLT by [10] for entire flags $(p^m, \dots, p^0)$ of nested subspaces $Q \supseteq p^m \supseteq \dots \supseteq p^0 = \{\mu\}$ which can be applied

(4) to *principal nested spheres* (PNS) by [11]
(5) and to the *intrinsic mean on the first geodesic principal component* by [9].

Currently we evaluate whether it can also be applied

(6) to the *barycentric subspaces* by [13] and
(7) to the *principal nested shape spaces* by [4].

3. Open Challenges

d) $\exists$ arbitrary smeariness on compact spaces?
e) Find conditions, when is $F \in \mathcal{C}^r$?
f) $\exists$ antismeariness $n^\gamma x_n = O_p(1)$ with $\gamma > 1/2$?

References

[1] Bhattacharya, R. and L. Lin (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proceedings of the American Mathematical Society 145*(1), 413–428.

[2] Bhattacharya, R. N. and V. Patrangenaru (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds I. *The Annals of Statistics 31*(1), 1–29.

[3] Bhattacharya, R. N. and V. Patrangenaru (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds II. *The Annals of Statistics 33*(3), 1225–1259.

[4] Dryden, I. L., K.-R. Kim, C. A. Laughton, and H. Le (2019). Principal nested shape space analysis of molecular dynamics data. *arXiv preprint arXiv:1903.09445*.

[5] Eltzner, B. and S. F. Huckemann (2018). A smeary central limit theorem for manifolds with application to high dimensional spheres. accepted (The Annals of Statistics), arXiv:1801.06581.

[6] Hotz, T. and S. Huckemann (2015). Intrinsic means on the circle: Uniqueness, locus and asymptotics. *Annals of the Institute of Statistical Mathematics 67*(1), 177–193.

[7] Huckemann, S. (2011). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *The Annals of Statistics 39*(2), 1098–1124.

[8] Huckemann, S., T. Hotz, and A. Munk (2010a). Intrinsic MANOVA for Riemannian manifolds with an application to Kendall's space of planar shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence 32*(4), 593–603.

[9] Huckemann, S., T. Hotz, and A. Munk (2010b). Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions (with discussion). *Statistica Sinica 20*(1), 1–100.

[10] Huckemann, S. F. and B. Eltzner (2018). Backward nested descriptors asymptotics with inference on stem cell differentiation. *The Annals of Statistics* (5), 1994 – 2019.

[11] Jung, S., I. L. Dryden, and J. S. Marron (2012). Analysis of principal nested spheres. *Biometrika 99*(3), 551–568.

[12] Mardia, K. V. and P. E. Jupp (2000). *Directional Statistics*. New York: Wiley.

[13] Pennec, X. (2018). Barycentric subspace analysis on manifolds. *The Annals of Statistics 46*(6A), 2711–2746.

[14] van der Vaart, A. (2000). *Asymptotic statistics*. Cambridge Univ. Press.

[15] Ziezold, H. (1977). Expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Transaction of the 7th Prague Conference on Information Theory, Statistical Decision Function and Random Processes A*, 591–602.

Super-Consistent Estimation of Points of Impact in Nonparametric Regression with Functional Predictors

Alois Kneip

(joint work with Dominik Poß, Dominik Liebl, Hedwig Eisenbarth, Tor D. Wager, Lisa Feldman Barrett)

Our methodology is motivated by data from a psychological experiment in which n participants were asked to continuously rate their emotional state while watching an affective video eliciting varying intensity of emotional reactions. This results in n random functions $X_i(t) \in \mathbb{R}$, with $t \in [a,b]$, where a denotes the start of the video and b the end. Psychologists are interested in understanding how a real response variable Y_i (overall ratings) relates to the fluctuations of the emotional states while watching the video, as this has implications for the way emotional states are assessed in research using such material. The problem is therefore to identify influential time points $\tau \in [a,b]$ with $X_i(\tau)$ possessing some significant impact on the response Y_i.

The general approach assumes an i.i.d. sample of data (X_i, Y_i), $i = 1, \ldots, n$, where $X_i = \{X_i(t), t \in [a,b]\}$ is a stochastic process with $\mathbb{E}(\int_a^b X_i(t)^2\, dt) < \infty$, $[a,b]$ is a compact subset of $\mathbb{R}$ and Y_i is a real valued random variable. It is assumed that the relationship between Y_i and the functional predictor X_i can be modeled as

$$Y_i = g\big(X_i(\tau_1), \ldots, X_i(\tau_S)\big) + \varepsilon_i,$$

where ε_i denotes the statistical error term with $\mathbb{E}(\varepsilon_i | X_i(t)) = 0$ for all $t \in [a,b]$. The number $0 \leq S < \infty$ and the points of impact $\tau_1, \ldots, \tau_S$ are unknown and have to be estimated from the data – without knowing the true model function g. The points of impact $\tau_1, \ldots, \tau_S$ indicate the locations at which the functional values

$X_i(\tau_1), \dots, X_i(\tau_S)$ have a specific influence on Y_i. Without loss of generality, we consider centered random functions X_i with $\mathbb{E}(X_i(t)) = 0$ for all $t \in [a, b]$.

We require that $g(x_1, \dots, x_S)$ is twice continuously differentiable, and that for all $r = 1, \dots, S$ the partial derivatives $\partial g(x_1, \dots, x_s)/\partial x_r$ are continuous almost everywhere as well as $0 < \vartheta_r := |\mathbb{E}(\frac{\partial}{\partial x_r} g(X_i(\tau_1), \dots, X_i(\tau_S)))| < \infty$.

Surprisingly, the unknown function g does not have to be estimated in order to estimate the points of impact $\tau_1, \dots, \tau_S$. Estimating points of impact, however, necessarily depends on the structure of X_i. Motivated by our application we consider Gaussian processes with rough sample paths such as (fractional) Brownian motion, Ornstein-Uhlenbeck processes, etc. The following assumption on the covariance function of X_i describes a very large class of such stochastic processes and allows us to derive precise theoretical results:

Assumption. For some open subset $\Omega \subset \mathbb{R}^3$ with $[a, b]^2 \times [0, b - a] \subset \Omega$, there exists a twice continuously differentiable function $\omega : \Omega \to \mathbb{R}$ as well as some $0 < \kappa < 2$ such that for all $s, t \in [a, b]$

$$\sigma(s, t) = \omega(s, t, |s - t|^\kappa).$$

Moreover, $0 < \inf_{t \in [a,b]} c(t)$, where $c(t) := -\frac{\partial}{\partial z} \omega(t, t, z)|_{z=0}$.

Under these conditions a generalization of Stein's lemma leads to

$$f_{XY}(s) := \mathbb{E}\big(X_i(s) Y_i\big) = \sum_{r=1}^{S} \vartheta_r \sigma(s, \tau_r) \quad \text{for all } s \in [a, b].$$

Since $\sigma(s, t)$ is not two times differentiable at $s = t$, the cross-covariance $f_{XY}(s)$ will not be two times differentiable at $s = \tau_r$, for all $r = 1, \dots, S$, resulting in kink-like features at τ_r. A natural strategy to estimate τ_r is to detect these kinks by considering the following modified central difference approximation of the second derivative of f at a point $s \in [a - \delta, b - \delta]$ for some $\delta > 0$. Defining the auxiliary process $Z_{\delta,i}(s) := X_i(s) - \frac{1}{2}(X_i(s - \delta) + X_i(s + \delta))$, we obtain

$$\mathbb{E}(Z_{\delta,i}(s) Y_i) = f_{XY}(s) - \frac{1}{2}(f_{XY}(s + \delta) + f_{XY}(s - \delta)).$$

Of course, $\mathbb{E}(Z_{\delta,i}(s) Y_i)$ is not known and we have to rely on $n^{-1} \sum_{i=1}^{n} Z_{\delta,i}(s) Y_i$ as its estimate. Under our setting we will have $\mathbb{V}(Z_{\delta,i}(s) Y_i) = O(\delta^\kappa)$, implying

$$\frac{1}{n} \sum_{i=1}^{n} Z_{\delta,i}(s) Y_i - \mathbb{E}(Z_{\delta,i}(s) Y_i) = O_P\left(\sqrt{\delta^\kappa / n}\right).$$

Estimates $\widehat{\tau}_1, \dots, \widehat{\tau}_{\widehat{S}}$ are then obtained by identifying the **local maxima** of $\frac{1}{n} \sum_{i=1}^{n} Z_{\delta,i}(s)$ which exceed a prespecified threshold $\lambda > 0$, i.e. $\frac{1}{n} \sum_{i=1}^{n} Z_{\delta,i}(\widehat{\tau}_j) > \lambda$ for all $j = 1, \dots, \hat{S}$.

A practical and asymptotically valid threshold specification which performed well in our simulation studies is given by $\lambda = A((\mathbb{E}(Y_i^4))^{1/2} \log\big((b - a)/\delta\big)/n)^{1/2}$, where $\mathbb{E}(Y_i^4)$ is estimated by $\widehat{\mathbb{E}}(Y_i^4) = n^{-1} \sum_{i=1}^{n} Y_i^4$ and $A = \sqrt{2\sqrt{3}}$. At the same time one may choose $\delta = 1/\sqrt{n}$. Under some additional moment conditions on Y_i it can then be shown that

$$\max_{r=1,\ldots,\widehat{S}} \min_{s=1,\ldots,S} |\widehat{\tau}_r - \tau_s| \;=\; O_P(n^{-1/\kappa}).$$

as well as

$$P(\widehat{S} = S) \;\to\; 1 \quad \text{as} \quad n \to \infty.$$

Given estimates $\widehat{\tau}_1, \ldots, \widehat{\tau}_{\widehat{S}}$, the function g may then be estimated by applying nonparametric regression methods, e.g. the Nadaraya-Watson kernel estimator.

A detailed description of the conceptual approach and asymptotic theory can be found in [1]. The paper also provides a detailed empirical study of data from a psychological experiment.

References

[1] Poß, D., Liebl, D., Kneip, A., Eisenbarth, H., Wager, T.D., and Feldman Barrett, L. (2019): Super-Consistent Estimation of Points of Impact in Nonparametric Regression with Functional Predictors; arXiv:1905.09021

Nonparametric Tolerance Tubes for Tracking Functional Data

Regina Liu
(joint work with Yi Fan)

Tolerance intervals and tolerance regions are important tools for process monitoring or statistical quality control of univariate and multivariate data, respectively. We discuss their generalization to tolerance tubes in the infinite dimensional setting for functional data. In addition to the generalizations of the commonly accepted definitions of the tolerance level of beta-content or beta-expectation, we introduce the new definition of alpha-exempt beta-expectation tolerance tube. The latter loosens the definition of beta-expectation tolerance tube by allowing alpha (pre-set using domain knowledge) portion of each functional be exempt from the requirement.

Those proposed tolerance tubes are completely nonparametric and broadly applicable. We discuss their general properties, and show that the alpha exempt beta-expectation tolerance tube is particularly useful in the setting where occasional short term aberrations of the functional data are deemed acceptable (or unpreventable) and they do not cause substantive deviation of the norm. This desirable property is elaborated further and illustrated with both simulations and real applications in continuous monitoring of blood glucose level in diabetes patients as well as of aviation risk patterns of aircraft landings.

A novel framework for the statistical analysis of Functions on Surfaces

Eardi Lila
(joint work with John Aston)

We establish a statistical framework for the analysis of Functions on Surfaces (FoSs) [1]. FoSs are geometric objects coupled with functional information, displaying both geometric and functional variations. In Figure 1, we show an example of FoSs representing brain geometries and associated cortical thickness maps. Such data are becoming increasingly common, in particular in the medical imaging community. Nonetheless, we are still lacking statistical approaches that can parsimoniously model geometric and functional aspects of these complex objects.

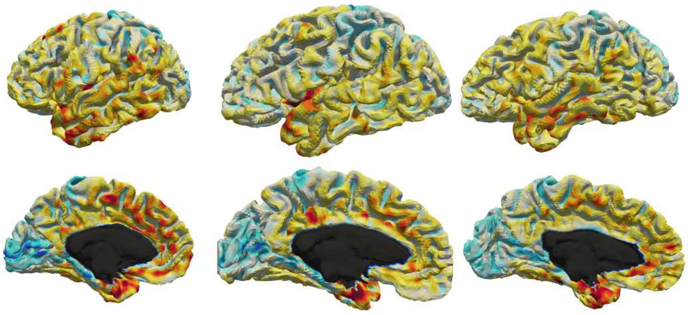

Figure 1. Surfaces representing the geometry of the brain's left hemispheres of three different subjects, with an associated scalar map representing the cerebral cortex thickness of the subjects.

We propose a general model for FoSs, where geometric variations are modeled as random diffeomorphic deformations of a template, while functional variations are modeled as random functions supported on the template. Diffeomorphic deformations are smooth functions that are invertible and have a smooth inverse. They have the desirable property of preserving the topology of the deformed objects and avoiding the formation of singularities. However, diffeomorphic functions belong to a non-Euclidean space, invalidating classical linear statistical approaches. We construct diffeomorphisms as flows of ordinary differential equations governed by time-varying smooth vector fields ([2]) and use the smooth vector fields to linearly represent the associated diffeomorphic functions. In order to quantify geometric and functional variations, we introduce estimators of the underlying unknown quantities within the proposed statistical model.

We apply the proposed model to the FoSs in Figure 1 to quantify the inherent variabilities of cortical thickness maps and brain geometries across subjects. We finally study the associations between variations in the brain shape and variations in the cortical thickness of the brain.

References

[1] Lila, Eardi and Aston, John A. D. (2019), Statistical Analysis of Functions on Surfaces, with an application to Medical Imaging, *Journal of the American Statistical Association*, in press.

[2] Younes, Laurent (2010), Shapes and diffeomorphisms, *Springer*.

Estimating functionals

László Györfi

In this talk I considered estimating three functionals: differential entropy, residual variance and Bayes error probability.

1. Differential entropy

Let X be a random vector taking values in $\mathbb{R}^d$ with probability density function $f(x)$, then its differential entropy is defined by

$$H(f) = -\int f(x) \ln f(x) dx. \tag{1}$$

Kozachenko and Leonenko [3] introduced the nearest neighbor entropy estimate as follows. Put $\rho_{n,i} = \min_{j\neq i, j\leq n} \|X_i - X_j\|$. Then the nearest neighbor entropy estimate is

$$H_n = \frac{1}{n}\sum_{i=1}^{n} \ln((n-1)\rho_{n,i}^d v_d) + C_E, \tag{2}$$

where C_E is the Euler-Mascheroni constant: $C_E = -\int_0^\infty e^{-t}\ln t dt = 0.5772...$ and v_d denotes the volume of the unit sphere in $\mathbb{R}^d$.

If f has bounded support and $\int f(x)\ln^2(f(x)+1)dx < \infty$, then

$$\mathbb{V}ar(H_n) = O(1/n). \tag{3}$$

Furthermore, if f is Lipschitz continuous and it has a bounded support, then

$$\mathbb{E}\{H_n\} - H(f) = O(n^{-1/d}). \tag{4}$$

These results can be applied for testing independence. Consider a sample of $\mathbb{R}^d \times \mathbb{R}^{d'}$-valued random vectors $(X_1, Y_1), \dots, (X_n, Y_n)$ with independent and identically distributed (i.i.d.) pairs. Assume that the distribution of (X, Y) has a density, which is denoted by f, while p and q stand for the densities of X and Y, respectively. We are interested in testing the null hypothesis that X and Y are independent, i.e.,

$$\mathcal{H}_0 : f = p \times q, \tag{5}$$

while making minimal assumptions regarding the densities.

If $H(f)$, $H(p)$ and $H(q)$ exist and are finite, then the null hypothesis (5) is equivalent to

$$KL(f, p \times q) = 0,$$

where KL denotes the Kullback-Leibler divergence. Because of

$$KL(f, p \times q) = H(p) + H(q) - H(f), \tag{6}$$

Berrett and Samworth [1] considered the test statistic

$$T_n = H_{n,p} + H_{n,q} - H_{n,f}, \tag{7}$$

where $H_{n,p}$, $H_{n,q}$ and $H_{n,f}$ are the Kozachenko-Leonenko entropy estimates of $H(p)$, $H(q)$ and $H(f)$, respectively.

Similarly to (3) and (4) we get that

$$\mathbb{V}ar(T_n) = O(1/n). \tag{8}$$

and

$$\mathbb{E}\{T_n\} - KL(f, p \times q) = O(n^{-1/d}). \tag{9}$$

Introduce the critical value $C_n = \omega_n(n^{-1/2} + n^{-1/d})$ with $\omega_n \to \infty$ such that $C_n \to 0$. Accept the null hypothesis of independence if

$$T_n \leq C_n,$$

and reject otherwise. Then, (8) and (9) imply that the error of the first and of the second kind tend to zero.

2. Residual variance

The residual variance is the smallest achievable mean-squared error in regression function estimation. For the d dimensional feature vector X and response variable Y, Devroye, Györfi, Lugosi and Walk [2] studied the problem of estimating the residual variance. The problem is equivalent to estimating the second moment of the regression function of Y on X. They introduced a nearest-neighbor-based estimate and obtained a normal limit law for the estimate when X has a density. Computed the asymptotic variance explicitly and derived a non-asymptotic bound on the variance that does not depend on the dimension d. The asymptotic variance does not depend on the smoothness of the density of X or on the regression function. Illustrated the use of the new estimate through testing whether a component of the vector X carries information for predicting Y.

3. Bayes error probability

For the d dimensional feature vector X and binary label Y, the Bayes error probability is the smallest achievable error probability in binary classification. The obvious way for estimating the Bayes error probability is the plug-in estimate, where from a training sample one creates a classification rule, and the estimate of the Bayes error probability is simply the empirical error of this classification rule calculated from a testing sample. The problem is

- either to show an estimate of the Bayes error probability with the rate of convergence better than that of the plug-in estimate,
- or to prove that it is impossible to construct an estimate with such fast rate of convergence.

References

[1] Berrett, T. B. and Samworth, R. J.: Nonparametric independence testing via mutual information. *Biometrika*, to appear, 2019.

[2] Devroye, L., Györfi, L., Lugosi, G. and Walk, H.: A nearest neighbor estimate of the residual variance, *Electronic Journal of Statistics*, 12:1752–1778, 2018.

[3] Kozachenko, L. F. and Leonenko, N. N.: Sample estimate of entropy of a random vector. *Problems of Information Transmission*, 23:95–101, 1987.

On the choice of suitable distances in Functional Data Analysis

Antonio Cuevas

(joint work with José R. Berrendero, Beatriz Bueno-Larraz,
Alejandro Cholaquidis)

This talk is concerned with Functional Data Analysis (FDA), i.e. with statistical problems (mostly related to classification or regression) in which the available data are functions or, to be more precise, the data are trajectories drawn from a stochastic process $\{X(t), \in [0,1]\}$. Unlike ordinary multivariate analysis (where the data live in $\mathbb{R}^d$), in FDA there is a strong case to consider the use of different distances between the sample (functional) data, according to the statistical methodology we are interested in. Of course, the basic point here is the obvious fact that there is no unique "natural" way of measuring the distance between two functions.

Whereas the standard L^2 distance (for square integrable functions) and the supremum distance (for continuous functions) are the most popular choices, there are some specific situations where the use of some other metrics makes sense. In this talk we will consider, from both a theoretical and a practical point of view, a few instances of these situations.

A) A functional Mahalanobis distance

In classical multivariate analysis, the Mahalanobis distance between two points x, y in the Euclidean space (with respect to a distribution with non-singular covariance matrix Σ) is defined by $M(x,y) = [(x-y)'\Sigma^{-1}(x-y)]^{1/2}$. Such distance is extremely useful in a number of applications, including classification an exploratory data analysis. Clearly, $M(x,y)$ is nothing but a "statistically meaningful" version of the Euclidean metric, aimed at taking into account the covariance structure of the data. A major hurdle for the definition of a functional version of $M(x,y)$ is the fact that a (functional) covariance operator is typically compact and, therefore, not invertible. A proposal aimed at overcoming such a problem (and still keeping the essential ideas and properties behind Mahalanobis metric) is given in [2]. The proposed definition relies on the use of a Tikhonov regularization method, combined with the natural distance in $\mathcal{H}(K)$, the Reproducing Kernel Hilbert Space (RKHS) associated with the covariance function $K = K(s,t)$ of the process generating the data.

B) An RKHS-based metric in functional regression

The RKHS metric is useful as well in order to define functional regression models with a scalar response. In particular, [3, 5] analyse the definition of the following functional regression model in RKHS-terms,

$$Y_i = \alpha_0 + \langle X_i, \beta \rangle_K + \varepsilon_i, \ i = 1, \ldots, n, \tag{1}$$

where β is the slope function $\beta \in \mathcal{H}(K)$ and $\langle \cdot, \cdot \rangle_K$ denotes here a suitable extension of the inner product in the RKHS $\mathcal{H}(K)$, defined in terms of the so-called Loève's isometry. Such an extension is needed since, typically, the trajectories of the process $X = X(t)$ do not belong (with probability one) to the corresponding RKHS, $\mathcal{H}(K)$.

A model of type (1) turns out to be particularly useful in order to define, and theoretically motivate, variable selection methods. The basic reason for this is the fact that all finite-dimensional models of type

$$Y_i = \alpha_0 + \sum_{i=1}^{p} \beta_i X(t_i), \ i = 1, \ldots, n, \tag{2}$$

(with $t_1, \ldots, t_p \in [0, 1]$, $\beta_1, \ldots, \beta_p \in \mathbb{R}$) appear just as particular cases of (1) for suitable choices of β.

Likewise, in the framework of logistic-functional regression, an RKHS-based model of type

$$\mathbb{P}(Y = 1 \,|\, X = x) \ = \ \frac{1}{1 + \exp\{-\beta_0 - \langle x, \beta \rangle_K\}},$$

presents a number of theoretical and practical advantages (when compared with its L^2-based counterpart) which are discussed in [2].

See also [4] for related ideas in the setting of functional classification.

C) A "visual" distance for functional data

Finally, a "visual" distance for functional data, essentially based on the Hausdorff metric between the corresponding hypographs, is explored in [6] in the framework of supervised classification with functional data. As discussed in that paper, the use of a Hausdorff-based distance is particularly appropriate in those situations (involving, e.g., mass spectra data) where one is concerned with the shape of the curves and one must take into-account their visual proximity, including possible small lateral shifts. The theoretical and practical aspects of this idea will be briefly summarized in the talk.

The mentioned Hausdorff-based metric (as well as other closely related ideas) has been also considered, from different points of view, in [7], [8] and [9].

References

[1] Berrendero, J.R., Bueno-Larraz, B. and Cuevas, A. (2018). On Mahalanobis distance in functional settings. *Manuscript, arXiv:1803.06550.*

[2] Berrendero, J.R., Bueno-Larraz, B. and Cuevas, A. (2018a). On functional logistic regression via RKHS's. *Manuscript, arXiv:1812.00721.*

[3] Berrendero, J.R., Bueno-Larraz, B. and Cuevas, A. (2019). An RKHS model for variable selection in functional linear regression. *Journal of Multivariate Analysis*, 170, 22-45.

[4] Berrendero, J.R., Cuevas, A. and Torrecilla, J.L. (2018b). On the use of reproducing kernel Hilbert spaces in functional classification. *Journal of the American Statistical Association*, 113, 1210-1218.

[5] Berrendero, J.R., Cholaquidis, A. and Cuevas, A. (2019a). On the estimation of the slope function in the RKHS-based functional regression model. *Manuscript in preparation.*

[6] Cholaquidis, A., Cuevas, A., and Fraiman, R. (2017). On visual distances for spectrum-type functional data. *Advances in Data Analysis and Classification*, 11, 5-24.

[7] Cuevas, A. and Fraiman, R. (1998). On visual distances in density estimation: The Hausdorff choice. *Statistics & Probability Letters*, 40, 333-341.

[8] Rockafellar, R.T. and Wets, R.J.B. (2009). *Variational Analysis.* Springer, New York.

[9] Sendov, B. (1990). *Hausdorff Approximations.* Kluwer, Dordrecht.

Scalar-on-function local linear regression and beyond

Frédéric Ferraty
(joint work with Stanislav Nagy)

When regressing nonparametrically a scalar response Y on any explanatory random function X, the common terminology refers to scalar-on-function nonparametric regression (see [6] for an overview on functional local constant regression). A natural development of functional local constant regression is the functional local linear regression (i.e local linear regression when the predictor is a random function). Actually, one can find in the literature only two papers dealing with functional local linear regression. The first published paper [3] proposes a projection approach but the asymptotics suffers from a lack of rigorousness and the second [4] is a pure theoretical work providing an alternative estimating procedure by regularizing a non bounded linear operator. Nevertheless, the scalar-on-function local linear regression is far to be popular as it is the case in the multivariate (i.e. non functional) case.

An exciting by-product of the functional local linear regression is its ability of providing an easy and fast way for estimating the functional derivative m'_x of the regression operator m at any function x which represents the local linear approximation of the regression operator m around x: for any v in a neighbourhood of x, it exists $\zeta = x + tv$ with $t \in (0,1)$ so that $m(x+v) = m(x) + \langle m'_x, v\rangle + \frac{1}{2}\langle m''_\zeta v, v\rangle$, where m''_ζ is a Hilbert-Schmidt linear operator. But why estimating the functional derivative of the regression operator? A first motivation is given in the pioneering works [7, 10] where estimating procedures are developed without considering the local linear regression setting. The authors demonstrated the usefulness of the concept of functional derivative for commenting results. As a continuation of these works and to understand how one can use the functional derivative in a simple way, let us go back to the functional Taylor expansion of the regression operator. For any positive real η small enough and any direction u (i.e. $\|u\| = 1$), one has $m(x + \eta u) - m(x) = \eta\langle m'_x, u\rangle + O(\eta^2)$. Then, the first order approximation for the range of the difference $m(x + \eta u) - m(x)$ belongs to the interval $[-\eta\|m'_x\|, \eta\|m'_x\|]$: smaller is $\|m'_x\|$ and less sensitive to small perturbations on x is $m(x)$. In some sense, $\|m'_x\|$ can be seen as a measure of reliability for the prediction of m at x. But the functional derivative may appear as a successful tool in very interesting statistical problems. For instance, consider the single functional index model $m(x) = \mu + g(\langle \beta, x \rangle)$ (see [2, 1, 5, 9]) where the scalar response interacts with the functional covariate only through a functional direction β combined with a real-valued link function g. Extending the ADE method introduced in [8] to the functional setting, it is easy to show that $E(m'_X)$ is proportional to the functional direction β where $\|\beta\| = 1$ for identifiability purpose. Then, given a sample $(X_1, Y_1), \dots, (X_n, Y_n)$, as soon as one is able to get the estimations $\widehat{m'_{X_1}}, \dots, \widehat{m'_{X_n}}$ of the functional derivatives $m'_{X_1}, \dots, m'_{X_n}$, one can compute $\widehat{E(m'_X)} := n^{-1} \sum_i \widehat{m'_{X_i}}$ so that $\widehat{E(m'_X)}/\|\widehat{E(m'_X)}\|$ is an estimator of the functional index β.

These different examples emphasize the major role that can play the functional derivative of the regression operator in important aspects of statistics: methodology, reliability and interpretation. This is why we provide a thorough and comprehensive study of the functional local linear regression to make it the benchmark method in the setting of scalar-on-function nonparametric regression. For both local linear estimators (regression operator and its functional derivative), original technical tools are developed for deriving their asymptotic behaviour. Easiness of implementation as well as nice finite sample properties of our local linear estimators are emphasized on simulated datasets and a benchmark growth dataset is used to demonstrate the important role that can play the functional derivatives.

References

[1] Ait-Saidi, A., Ferraty, F., Kassa, R. and Vieu, P. (2008). Cross-validated estimation in the single-functional index model. *Statistics* **42** 475-494.

[2] Amato, U., Antoniadis, A. and De Feis I. (2006). Dimension reduction in functional regression with application. *Comput. Statist. Data Anal.* **50** 2422-2446.

[3] Baíllo, A., Grané, A. (2009). Local linear regression for functional predictor and scalar response. *J. Multivariate Anal.* **100** 102-111.

[4] Berlinet, A., Elamine, A., Mas, A. (2011). Local linear regression for functional data. *Ann. Inst. Statist. Math.* 63 1047-1075.

[5] Chen, D., Hall, P. and Müller, H-G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* **39** 1720-1747.

[6] Ferraty, F., Vieu, P. (2006). *Nonparametric functional data analysis.* Springer, New York.

[7] Hall, P., Müller, H.-G. and Yao, F. (2009). Estimation of functional derivatives. *Ann. Statist.* **37** 3307-3329.

[8] Härdle, W. and Stoker, T. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84** 986-995

[9] Jiang, C.-R., J.-L. Wang (2011). Functional single index models for longitudinal data. *Ann. Statist.* **39** 362-388.

[10] Müller, H.-G., Yao, F. (2010). Additive modelling of functional gradients. *Biometrika* **97** 791-805.

Bootstrapping linear spectral statistics of high-dimensional sample covariance matrices

Angelika Rohde
(joint work with Holger Dette)

Let $Y_1, \ldots, Y_n$ be independent, identically distributed p-dimensional centered random vectors with covariance matrix Σ_n and corresponding sample covariance matrix

$$\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i Y_i'.$$

With $\hat{\lambda}_{1,n}, \ldots, \hat{\lambda}_{p,n}$ denoting its eigenvalues, many important statistics of $\hat{\Sigma}_n$ can be written as linear spectral statistics

$$\hat{T}_n(f) = \sum_{j=1}^{p} f\big(\hat{\lambda}_{j,n}\big)$$

for suitably regular real-valued functions f. For instance, the trace $\mathrm{tr}(\hat{\Sigma}_n)$ is the linear spectral statistics $\hat{T}_n(f)$ with f equals the identity map and the log-determinant $\log\det\hat{\Sigma}_n$ equals $\hat{T}_n(\log)$. In the high-dimensional scenario $p/n \to c > 0$, the centered version $\hat{T}_n(f) - \mathbb{E}\hat{T}_n(f)$ has a non-degenerate distribution which depends in an intricate way on the distribution $\mathcal{L}(Y_1)$ of Y_1. Results on the natural question of a nonparametric bootstrap approximation are negative, however ([2]). While the classical sampling with replacement bootstrap already fails for $f(x) = x^2$, the traditionally more robust 'm out of n' procedure does not even preserve the limiting ratio c of dimension and sample size if $m \ll n$. The latter ratio c, however, appears already explicitly in the existing limiting distribution in the simplest case where f is the identity and $Y_1, \ldots, Y_n$ are standard normally distributed, see e.g. [1]. Here, we provide a powerful and fully nonparametric bootstrap approximation of linear spectral statistics in this high-dimensional context. The idea is to rely on an 'm out of n' procedure ($m = o(n)$) while suitably subsampling 'q out of p' dimensions in order to keep the ratio

$$\frac{p}{n} = \frac{q}{m}$$

the same. Our new approach is based on the crucial observation that in most situations of interest, a subvector $Y_{1,sub}$ of Y_1, selected according to an appropriate random sampling mechanism, provides a covariance matrix with similar spectral distribution as the full vector Y_1 (see assumption (A4) below).

The new bootstrap algorithm

(i) For $m = o(\sqrt{n})$, draw an iid sample $Y_1^*, \ldots, Y_m^*$ from the empirical distribution

$$\hat{\mathbb{P}}_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{Y_k}.$$

(ii) Define the bootstrap sample

$$Z_i^* = (Y_{ij_1}^*, \dots, Y_{ij_q}^*)', \quad i = 1, \dots, m,$$

using the coordinates $j_1, \dots, j_q$ selected 'appropriately'.

For any Hermitian matrix $A \in \mathbb{R}^{p \times p}$, $\lambda_1(A), \dots, \lambda_p(A)$ denote its (possibly multiple) eigenvalues and

$$\mu^A = \frac{1}{p} \sum_{k=1}^{p} \delta_{\lambda_k(A)}$$

its spectral measure. Corresponding to the bootstrap sample covariance matrix

$$\hat{\Sigma}_n^* = \frac{1}{m} \sum_{i=1}^{m} Z_i^* {Z_i^*}',$$

$\mu^{\hat{\Sigma}_n^*}$ denotes its spectral measure and

$$\hat{T}_n^*(f) = q \int f \mathrm{d}\mu^{\hat{\Sigma}_n^*}$$

the bootstrapped linear spectral statistics.

Assumptions.

(A1) $p/n \to c > 0$ (High-dimensionality)

(A2) The sequence $\left(\mu^{\Sigma_n}\right)_{n \in \mathbb{N}}$ of spectral distributions is tight.

(A3) The p-dimensional random vector $Y_1^{(n)}$ is assumed to be of the form

$$Y_1^{(n)} = A_n X_1, \tag{1}$$

where $X_1 = (X_{11}, X_{12}, \dots)'$ is an infinite-dimensional random vector with iid entries, satisfying

- $\mathbb{E}X_{11} = 0$
- $\mathbb{E}X_{11}^2 = 1$
- $\mathbb{E}X_{11}^4 = 3$,

and the $(p \times \infty)$-matrix A_n has rows in ℓ^2 such that $A_n A_n' = \Sigma_n \in \mathbb{R}^{p \times p}$.

(A4) (Representative subpopulation condition)
The sequence of p_n-dimensional vectors $(Y_n)_{n \in \mathbb{N}}$ is assumed to possess the following properties.

(i) There exists a possibly random strategy of selecting a sequence of q_n-dimensional subvectors $Y_{i,sub}$ with corresponding covariance matrices $\tilde{\Sigma}_{q_n}$ such that

$$\mu^{\tilde{\Sigma}_{q_n}} - \mu^{\Sigma_n} \Rightarrow 0 \quad \text{as } q, n \to \infty \quad \text{in probability.}$$

(ii) If $\Pi_{p_n q_n}$ denotes the projection corresponding to the possibly random selection strategy, that is $Y_{n,sub} = \Pi_{p_n q_n} Y_n$, then there exists for almost all realizations a decomposition of the form

$$\Pi_{p_n q_n} A_n = L_n + R_n$$

where the matrix L_n has at most $\mathcal{O}(q_n)$ non-zero columns and $||R_n||_{S_2}^2 = o(1)$.

Assumption (A1) and (A2) are classical assumptions for the CLT of linear spectral statistics. There, instead of (A3), the more restrictive representation

$$Y_1 = \Sigma_n^{1/2} X_n$$

is required, where X_n is p-dimensional with iid coordinates and the same moment conditions as in (A3) (the assumption on the fourth moment can be relaxed – in the classical context as well as in our bootstrap context). Our crucial requirement and innovation is condition (A4). It is satisfied in particular if the entries of Y_1 are p consecutive elements of a stationary process under mild regularity assumptions.

Our main result is as follows. Here, $\Rightarrow$ stands for weak convergence, d_{BL} denotes the dual bounded Lipschitz metric and as usual, $\rightarrow_{\mathbb{P}}$ refers to convergence in probability.

Theorem 1 (Bootstrap consistency)**.** *Grant assumptions (A1) – (A4). Then*

$$\mu^{\hat{\Sigma}_n} - \mu^{\hat{\Sigma}_n^*} \Longrightarrow 0$$

in probability and

$$d_{BL}\Big[\mathcal{L}\big(\hat{T}_n^*(f) - \mathbb{E}^*\hat{T}_n^*(f)\,\big|\,Y_1,\ldots,Y_n\big), \mathcal{L}\big(\hat{T}_n(f) - \mathbb{E}\hat{T}_n(f)\big)\Big] \longrightarrow_{\mathbb{P}} 0.$$

References

[1] Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large dimensional sample covariance matrices. *Annals of Probability* **32**, 553–605.

[2] El Karoui, N. and Purdom, E. (2016). The bootstrap, covariance matrices and PCA in moderate and high-dimensions. *arXiv:1608.00948.*

Reporter: Yoav Zemel

Participants

Prof. Dr. Ying Chen
Department of Mathematics
National University of Singapore
Lower Kent Ridge Road
Singapore 119260
SINGAPORE

Prof. Dr. Antonio Cuevas
Departamento de Matemáticas
Facultad de Ciencias
Universidad Autónoma de Madrid
Ciudad Universitaria de Cantoblanco
28049 Madrid
SPAIN

Prof. Dr. Aurore Delaigle
School of Mathematics and Statistics
The University of Melbourne
Parkville, VIC 3010
AUSTRALIA

Prof. Dr. Holger Dette
Fakultät für Mathematik
Ruhr-Universität Bochum
44780 Bochum
GERMANY

Dr. Ian Dryden
School of Mathematical Sciences
The University of Nottingham
University Park
Nottingham NG7 2RD
UNITED KINGDOM

Prof. Dr. Brittany T. Fasy
School of Computing
Montana State University
363 Barnard Hall
Bozeman, MT 59717
UNITED STATES

Prof. Dr. Frédéric Ferraty
Institut de Mathématiques de Toulouse
Université Paul Sabatier
118, route de Narbonne
31062 Toulouse Cedex 9
FRANCE

Prof. Dr. Laszlo Györfi
Department of Computer Science and
Information Theory
Budapest University of Technology
and Economics
Stoczek u. 2
1521 Budapest
HUNGARY

Prof. Dr. Wolfgang K. Härdle
Wirtschaftswissenschaftliche Fakultät
Ladislaus v. Bortkiewicz Chair of
Statistics
Humboldt-Universität zu Berlin
Unter den Linden 6
100117 Berlin
GERMANY

Prof. Dr. Stephan Huckemann
Institut für Mathematische Stochastik
Georg-August-Universität Göttingen
Goldschmidtstrasse 7
37077 Göttingen
GERMANY

Prof. Dr. Moritz Jirak
Institut für Mathematische Stochastik
Technische Universität Braunschweig
Postfach 3329
38023 Braunschweig
GERMANY

Prof. Dr. Alois R. Kneip
Fachbereich Wirtschaftswissenschaften
Hausdorff Center for Mathematics
Endenicher Allee 60
53115 Bonn
GERMANY

Eardi Lila
Department of Mathematics and Statistics
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0WB
UNITED KINGDOM

Prof. Dr. Regina Y. Liu
Department of Statistics
Rutgers University
501 Hill Center
110 Frelinghuysen Road
Piscataway, NJ 08854-8019
UNITED STATES

Prof. Dr. James Stephen Marron
Department of Statistics and Operations Research
University of North Carolina
Chapel Hill, NC 27599-3260
UNITED STATES

Prof. Dr. Alexander Meister
Institut für Mathematik
Lehrstuhl für Mathematische Statistik
Universität Rostock
18051 Rostock
GERMANY

Prof. Dr. Hans-Georg Müller
Department of Statistics
University of California
469 Kerr Hall
Davis, CA 95616-8705
UNITED STATES

Prof. Dr. Victor M. Panaretos
Institut de Mathématiques
École Polytechnique Fédérale de Lausanne
Station 8
1015 Lausanne
SWITZERLAND

Prof. Dr. Angelika Rohde
Fakultät für Mathematik
Albert-Ludwigs-Universität Freiburg
LST für Stochastik
Ernst-Zermelo-Strasse 1
79104 Freiburg i. Br.
GERMANY

Dr. Shahin Tavakoli
Department of Statistics
University of Warwick
MB2.14 Mathematical Sciences Building
Coventry CV4 7AL
UNITED KINGDOM

Dr. Katharine Turner
Mathematical Sciences Institute
Australian National University
Union Lane
Acton ACT 2601
AUSTRALIA

Prof. Dr. Jane-Ling Wang
Department of Statistics
University of California
469 Kerr Hall
Davis, CA 95616-8705
UNITED STATES

Dr. Yoav Zemel
Mathematisches Institut
Georg-August-Universität Göttingen
Bunsenstrasse 3-5
37073 Göttingen
GERMANY