ABSTRACTS

"A Condensation of Skew-Symmetric Distributions in $< 60'$"

Adelchi Azzalini
University of Padua, Italy
azzalini@stat.unipd.it

The aim is to provide a broad introduction to the so-called area of skew-symmetric probability distributions, targeted to people without or with limited knowledge of this theme. Some general aspects will be presented, with emphasis on motivation and qualitative aspects more than formalism. Special noteworthy cases will be presented, to illustrate the potential of these tools in practical problems. We close with a brief sketch of an associated R package.

"Part I: Functional Boxplots for Visualization of Complex Curve/Image Data: An Application to Precipitation and Climate Model Output"

Marc G. Genton
CEMSE Division, KAUST, Saudi Arabia
stsda.kaust.edu.sa

In many statistical experiments, the observations are functions by nature, such as temporal curves or spatial surfaces/images, where the basic unit of information is the entire observed function rather than a string of numbers. For example the temporal evolution of several cells, the intensity of medical images of the brain from MRI, the spatio-temporal records of precipitation in the U.S., or the output from climate models, are such complex data structures. Our interest lies in the visualization of such data and the detection of outliers. With this goal in mind, we have defined functional boxplots and surface boxplots based on the center outwards ordering induced by band depth for functional data or surface data. We illustrate the construction of a functional boxplot on a series of sea surface temperatures related to the El Nino phenomenon and its outlier detection performance is explored by simulations. As applications, the functional boxplot is demonstrated on spatio-temporal U.S. precipitation data for nine climatic regions and on climate general circulation model (GCM) output. Further adjustments of the functional boxplot for outlier detection in spatio-temporal data are discussed as well. The talk is based on joint work with Ying Sun.

'Part II: Paradigms for a 21st Century University'"

KAUST is a graduate research institution, founded in 2009 by King Abdullah of Saudi Arabia, with research thrusts in energy, environment, food, and water for a sustainable planet, and supporting thrusts in core capabilities (modeling, simulation, analytics, software, and hardware). As a 45-sq km international academic village on the shores of the Red Sea, created to be a 21st century "House of Wisdom" in the tradition of the ninth century Bayt al Hikmah that gave the world some of its modern mathematics, physics, chemistry, and medicine, KAUST has been endowed with world- class facilities and has recruited a world-competitive research faculty. KAUST awards degrees in Applied Mathematics and Computational Science, Bioscience, Chemical Science, Chemical and Biological Engineering, Computer Science, Earth Science and Engineering, Electrical Engineering, Environmental Science and Engineering, Marine Science, Materials Science and Engineering, and Mechanical Engineering. Currently, KAUST enrolls about 800 students from about 60 different countries. The language of instruction is English. KAUST is co-educational and is established upon principles of intellectual freedom, non-discrimination, and merit-based promotion. For Fall 2013, KAUST seeks ambitious, academically talented, and highly motivated doctoral and master's candidates in sustainable technologies and the enabling sciences from the worlds leading institutions to participate in discovery and translation into start-up enterprises. The speaker will present KAUST's programs and take questions about life in today's Middle East and about starting a university from scratch.

"The Multivariate Extended Normal-Gamma Distribution"

Chris J. Adcock
Sheffield University Management School, UK
c.j.adcock@shef.ac.uk

This paper describes an extended version of the multivariate normal-gamma distribution and its properties. The distribution is closed under conditioning and exhibits the general property that it is possible for standardised values of skewness and kurtosis to become arbitrarily large. These properties imply that the model has potential for applications in financial economics, particularly for asset classes whose returns are severely asymmetric. All moments of the distribution exist; a property which is important for general portfolio selection. The multivariate extended skew-normal distribution arises as a special case.

"Skewed Mixtures in Model-Based Classification"

Paul D. McNicholas
University of Guelph, Canada
pmcnicho@uoguelph.ca

Over the past 50 years, finite mixture models have become a popular tool for classification and clustering. Finite mixture models are based on the underlying assumption that a population is a convex combination of a finite number of densities. Until quite recently, almost all work on clustering and classification using mixture models had been based on Gaussian mixture models. This includes work by Banfield and Raftery (1993), Celeux and Govaert (1995), Ghahramani and Hinton (1997), Tipping and Bishop (1999), McLachlan and Peel (2000), and Fraley and Raftery (2002), amongst others. Early work on clustering using mixtures of multivariate t-distributions [McLachlan and Peel (1998) and Peel and McLachlan (2000)] was the forerunner to several papers on clustering using mixtures of multivariate t-distributions, including those by McLachlan et al. (2007), Andrews and McNicholas (2011), and Baek and McLachlan (2011). Although mixtures of multivariate t-distributions have some advantages over their Gaussian counterparts, the component density is still symmetric. More recently, there has been a move towards using asymmetric densities for mixture model-based classification and clustering. The movement towards classification using non-elliptical densities will be discussed, including some very recent developments. Real and simulated data will be used to illustrate some of these approaches.

Andrews, J. L. and P. D. McNicholas (2011). Mixtures of modified t-factor analyzers for model-based clustering, classification, and discriminant analysis, *Journal of Statistical Planning and Inference* **141**, 1479–1486.

Baek, J. and G. J. McLachlan (2011). Mixtures of common t-factor analyzers for clustering high-dimensional microarray data, *Bioinformatics* **27**, 1269–1276.

Banfield, J. D. and A. E. Raftery (1993). Model-based Gaussian and non-Gaussian clustering, *Biometrics* **49**, 803–821.

Celeux, G. and G. Govaert (1995). Gaussian parsimonious clustering models, *Pattern Recognition* **28**, 781–793.

Fraley, C. and A. E. Raftery (2002). Model-based clustering, discriminant analysis, and density estimation, *Journal of the American Statistical Association* **97**, 611–631.

Ghahramani, Z. and G. E. Hinton (1997). The EM algorithm for factor analyzers, *Technical Report CRG-TR-96-1*, University of Toronto, Toronto.

McLachlan, G. J., R. W. Bean, and L. B.-T. Jones (2007). Extension of the mixture of factor analyzers model to incorporate the multivariate t-distribution, *Computational Statistics & Data Analysis* **51**, 5327–5338.

McLachlan, G. J. and D. Peel (1998). Robust cluster analysis via mixtures of multivariate t-distributions, In *Lecture Notes in Computer Science*, Volume 1451, pp. 658–666. Berlin: Springer-Verlag.

McLachlan, G. J. and D. Peel (2000). Mixtures of factor analyzers, In *Proceedings of the Seventh International Conference on Machine Learning*, San Francisco, pp. 599–606. Morgan Kaufmann.

Peel, D. and G. J. McLachlan (2000). Robust mixture modelling using the t distribution, *Statistics and Computing* **10**, 339–348.

Tipping, T. E. and C. M. Bishop (1999). Mixtures of probabilistic principal component analysers, *Neural Computation* **11**, 443–482.

"Fernando de Helguero: Abnormal Statistics Arising from Selection"

Giuliana Regoli
University of Peruggia, Italy
regoli@dipmat.unipg.it

The current literature on so-called skew-symmetric distributions is closely linked to the idea of a selection mechanism operated by some latent variable. We illustrate the pioneering work of Fernando de Helguero who in 1908 put forward a formulation for the genesis of non-normal distributions based on a selection mechanism which perturbs a normal distribution, via an argument which has close connections with the construction now widely used in this context. Arguably, de Helguero can then be considered the precursor of the current idea of skew-symmetric distributions. Unfortunately, a tragic quirk of fate did not allow him to pursue his project beyond the initial formulation and his work went unnoticed for the rest of the 20th century.

"Fitting Age-Specific Fertility Rates by a Flexible Generalized Skew-Normal Probability Density Function"

Bruno Scarpa
University of Padua, Italy
scarpa@stat.unipd.it

Mixture probability density functions have recently been proposed to describe some fertility patterns characterized by a bimodal shape. These functions are adequate when the fertility pattern is actually bimodal, but appear to be less useful when the shape of age-specific fertility rates is unimodal. A further model is proposed, based on flexible skew-symmetric probability density functions. This model is able to fit symmetric and skew patterns as well as to reflect humps observed in some fertility patterns. It is both more parsimonious than mixture distributions and more flexible, showing a good fit with several shapes (bimodal or unimodal) of fertility patterns. Empirical evaluation of the proposed model and comparisons with other functions to the Italian data, from 1952 to 2003 and US data from 1933 to 2006 are also discussed.

"A Parametric Test for Trend"

Tao Tan
McMaster University, Canada
tant@math.mcmaster.ca

When researchers work on time series or sequence, certain fundamental questions will naturally arise. One of them we are interested in is whether the series or sequence exhibits a gradual trend over time. In this talk, the main results of my master's thesis will be discussed. We propose a test statistic based on moving order statistics and establish the strict procedure to test for the presence of monotonic trends. We find that the test statistic under the null hypothesis that there is no trend follows the closed skew normal distribution. An efficient algorithm is developed to generate realizations from the distribution. A simulation study is conducted to evaluate the proposed test under the alternative hypotheses with linear, logarithmic and quadratic trend functions. Finally, a real data example is provided to illustrate the proposed procedure.
