WORKSHOP DOFIN

(DOnnées Fonctionnelles et géométrie de l'INformation)

Friday 26th January 2018

This event is organised by ENAC, the National School of Civil Aviation, in collaboration with AMIES, the Agency for Interaction in Mathematics with Business and Society. This workshop is the opportunity to present the results of the PEPS1-DOFIN project in collaboration between ENAC and the start-up Safety Line, based in Paris. This workshop aims also to open the discussion and exchanges between functional data statistics, information geometry and applied research.

Date and location: Friday, January 26, Amphi Bellonte, 8h00-16h45

ENAC, 6 Avenue Edouard Belin, 31400 – Toulouse, France.

Registration: registration to attend the workshop is now open until Tuesday 23th January 2018. Please note that the registration is free but mandatory (florence.nicol@enac.fr).

Contact: Please contact to Florence Nicol at florence.nicol@enac.fr

Invited speakers: Pr. Aloïs Kneip (Bonn University, Germany), Dr. Antonio Gattone (Chieti Pescara University, Italy), Dr. Cindie Andrieu and Dr. Baptiste Gregorutti (Safety Line, Paris).

Program:

8h00-08h30 Welcome - Opening session.

8h30-10h00 *Functional Data Statistics*, A. Kneip (Bonn University, Germany).

10h00-10h30 Coffee break.

10h30-12h00 *Introduction to information geometry and shape spaces*, S. Puechmorel and F. Nicol (ENAC, France).

12h00-13h30 Lunch break.

13h30-14h30 *Information geometry and shape clustering*, S.A. Gattone (Chieti Pescara University, Italy).

14h30-15h30 *Detection of bad runway conditions using radar landing tracks*, B. Gregorutti and C. Andrieu (Safety Line, France).

15h30-15h45 Coffee break.

15h45-16h45 Panel discussion, Invited speakers+Guests (TBD).

Abstracts:

- *Functional Data Statistics*, A. Kneip (Bonn University, Germany).

Functional data analysis is a very active field of research in contemporary mathematical statistics. The research in this area is motivated by many important applications in biometrics, econometrics and finance. In times of ever-increasing availability of large datasets, available observations can often be interpreted as densely sampled data that are generated by some underlying stochastic process in continuous time. It is usually assumed that the underlying data consist of smooth random functions taking values in some suitable Hilbert space. Regularization techniques and nonparametric smoothing procedures are usually applied to determine efficient estimation and inference procedures. A major challenge for future research is to overcome the limitations of standard Hilbert space analysis. Some important applications lead to samples of density functions, which lie on a nonlinear subspace of L^2. Experiments in biomedicine frequently yield functional data possessing a common structure in terms of typical successions of peaks and valleys. In all these cases the function spaces to be analyzed are not Hilbert spaces, and efficient statistical analysis has to be based on suitably defined metrics and/or sophisticated data transformations. In this context crucial theoretical and methodological questions remain to be resolved.

- *Introduction to information geometry and shape spaces*, S. Puechmorel and F. Nicol (ENAC, France).

Information geometry studies the geometrical structure of manifolds of probability distributions. The Fisher information provides such manifolds with a Riemannian metric that encodes some important statistical properties. While asymptotic theory of inference is related to the tangent spaces to the model manifold, affine connections allows global properties to be addressed. Higher order asymptotics are obtained the same way. While in Riemannian geometry the Levi-Civita connection plays a central role, it is not true for information geometry. Introduced by Amari (1980), the family of alpha-connections is the key ingredient to link geometry and statistics. The purpose of the talk is to introduce the basic concepts of information geometry, and to give a different view at alpha-connections, that generalizes the original definition. In a second part, the notion of shape space will be introduced. While not directly related to information geometry, it gives a mean of dealing with functional data enforcing invariance properties. Rooted in the pioneering work of Mumford and Michor, a shape space is obtained from a manifold of immersions as a quotient by a group of re-parametrization. This framework allows computation of distances between shapes, and offers a convenient setting for classical clustering algorithms.

- *Information geometry and shape clustering*, S.A. Gattone (Chieti Pescara University, Italy).

Shape analysis is a timely and interesting research field. The applications range from structural biology, computer vision, medical imaging to archaeology. We focus on the selection of an appropriate measurement of distance among observations with the aim of obtaining an unsupervised classification of shapes. Under the hypothesis that it is possible to extract from the shape a finite number of representing points, called landmarks, each landmark is modeled with a bivariate Gaussian distribution in which averages are geometric coordinates and variances take into account the shapes variability and measurement errors. The Fisher-Rao metric can be used as a Riemannian metric of the statistical model thus identified. Besides it is the only one compatible with the Fisher information. Consequently, geodesic paths can be computed as locally minimizing distances in the Fisher information sense. The methodology enables us to perform various type of analysis including shape comparisons, interpolation between observed shapes and shape predictions. Indeed the length of the geodesic path connecting two shapes can be used for quantifying shape differences and used in a cluster analysis of shapes.

- *Detection of bad runway conditions using radar landing tracks*, B. Gregorutti andC. Andrieu (Safety Line, France).

In air transportation, a huge amount of data is continuously recorded such as radar tracks that may be used for improving flight as well as airport safety. However, all known statistical algorithms, even those based on functional data, are unable to distinguish between a safety critical flight and another one departing from standard behavior, but otherwise safe. It is the case in airport safety when radar measurements are used for detecting incidents on airport surface. In this talk, we propose a change of paradigm by switching from a functional data framework to a geometrical one by representing curves as points in a shape manifold. In this way, any intrinsic structure of the data that is amenable to geometry can be directly encoded in the representation space. Based on an extension of a classical distance between shapes, a new one is defined, that explicitly takes into account the second derivative and can be related to slippery. Its properties are investigated in a first part, then some results on datasets of synthetic and real trajectories are presented.

PLAN de L'ENAC