Depth measures for the study of real and simulated ECG signals

Francesca Ieva and Anna Maria Paganoni

Abstract In biomedical context examples of functional data to be studied with suitable statistical techniques are more and more frequently appearing. The paper is focused on the analysis of ECG signals, both arising from real clinical practice and numerical simulations, as multivariate functional data. The main focuses of the analysis are the application to such data of new techniques arising from the generalization of depth measures theory, as well as their use for the validation of simulated signals.

Key words: Depth measures; Multivariate Functional Data; ECG signals

1 Introduction and Motivations

A natural tool to analyze functional data, when the interest is in pointing out an order among curves, is the idea of statistical depth. The depth provides a measure of centrality of an observation with respect to a given dataset or a population distribution. A generalization of the multivariate concept of depth measure [5] to functional data is given in [6]. In this work we deal with multivariate functional observations, i.e., statistical units where each component is a curve. Generalizations of the concept of depth for functional data to the multivariate functional case, as well as the use of depth measure for generalizing nonparametric tests [4] to the multivariate functional case can be found in [1]. In that work we propose, analyze, and apply a new concept of index of depth for multivariate functional data, obtained as an average of univariate centrality measures for univariate functional data. In [2] the construction of the functional boxplots [7] is extended to the more complex context of multivariate functional data. All these tools may be naturally applied in the biomedical context,

Francesca Ieva, Anna Maria Paganoni
Politecnico di Milano, p.zza Leonardo da Vinci 32, 20133 Milano, e-mail: francesca.ieva@mail.polimi.it, anna.paganoni@polimi.it

and in particular in those applications that deal with cardiovascular diseases diagnoses carried out using Electrocardiographic (ECG) devices. In fact, ECG signals can be considered as multivariate functional data with dependent components (see, among others, [3]). In the context of the analysis of real ECG signals, some issues of interest are, for example, classification of groups of curves with similar morphological patterns, multivariate functional outliers detection within a homogeneous group and classical inference on mean, quantiles and variance of a specified subpopulations. From a clinical point of view, the first issue concerns how to carry out a semi automatic diagnosis based only on the morphological deviations from physiological patterns induced by the presence of the disease of interest; the second one leads to profile "representative" curves for each pathology; finally the third one allows for the investigation of the presence of statistically significant differences between the subpopulations of pathological units and the physiological ones.

We will focus now on the multivariate functional outliers detection performed through suitable multivariate functional depth indexes and applied to the validation of ECG signals arising from numerical simulations.

2 Depth Measures for Multivariate Functional Data

In [1], starting from the definition of depth measure given in [6], a new concept of multivariate functional depth measure is presented and statistical properties are established and proved. Moreover, the modified version of the band depth is given, and it is used to implement the modified version of the functional boxplot, introduced in [2]. Let X a stochastic process with law P taking values on the space $\mathscr{C}(I)$ of real continuous functions on the compact interval I. The graph of a function $f \in \mathscr{C}(I)$ is the subset of the plane $G(f) = \{(t, f(t)) : t \in I\}$. The random band depth, of order $J \geq 2$, for a function $f \in \mathscr{C}(I)$ is then $BD^J_{P_X}(f) = \sum_{j=2}^{J} P_X\{G(f) \subset B(X_1, X_2, ..., X_j)\}$, where $B(X_1, X_2, ..., X_j)$, for $j = 2, ..., J$ is the random band in $\mathbb{R}^2$ delimited by $X_1, ..., X_j$, independent copies of the stochastic process X, defined as $B(X_1, ..., X_j) = \{(t, y(t)) : t \in I, \min_{r=1,...,j} X_r(t) \leq y(t) \leq \max_{r=1,...,j} X_r(t)\}$. In [1] a new definition of a band depth measure for multivariate functional data is given, i.e., data generated by a stochastic process $\mathbf{X}$ taking values in the space $\mathscr{C}(I; \mathbb{R}^s)$ of continuous functions $\mathbf{f} = (f_1, ..., f_s) : I \to \mathbb{R}^s$. A *multivariate band depth measure* is defined as

$$BD^J_{P_{\mathbf{X}}}(\mathbf{f}) = \sum_{k=1}^{s} p_k BD^J_{P_{X_k}}(f_k), \qquad p_k > 0 \ for\ k = 1, ..., s, \qquad \sum_{k=1}^{s} p_k = 1. \tag{1}$$

and proofs of the basic statistical properties of the multivariate band depth measure are provided in [1].

If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent copies of the stochastic process $\mathbf{X}$, the sample version of (1) can be introduced in order to conduct descriptive and inferential statistical analyses on a set of multivariate functional data $\mathbf{f}_1, ..., \mathbf{f}_n$ generated by the process $\mathbf{X}$. For any $\mathbf{f}$ in the sample $\mathbf{f}_1,\ldots,\mathbf{f}_n$ we can compute the depth as $BD_n^J(\mathbf{f}) = \sum_{k=1}^{s} p_k BD_{n,k}^J(f_k)$, where, for the function $f_k \in \mathscr{C}(I)$,

$$BD_{n,k}^J(f_k) = \sum_{j=2}^{J} \binom{n}{j}^{-1} \sum_{1\le i_1<i_2<\cdots<i_j\le n} \mathbb{I}\{G(f_k) \subset B(f_{i_1;k},\ldots f_{i_j;k})\}$$

and $\mathbb{I}\{G(f_k) \subset B(f_{i_1;k},\ldots,f_{i_j;k})\}$ indicates if the band determined by $(f_{i_1;k},\ldots,f_{i_j;k})$ contains the whole graph of f_k. The k component of the vector $\mathbf{f_i}$ is denoted by $f_{i;k}$. As proposed in [6] also in this multivariate functional setting we can move to the analogous of the modified band depth:

$$MBD_n^J(\mathbf{f}) = \sum_{k=1}^{s} p_k MBD_{n,k}^J(f_k), \tag{2}$$

where for the function $f_k \in \mathscr{C}(I)$ the modified band depth measures the proportion of time that the curve f_k is in the band, i.e., $MBD_{n,k}^J(f_k) = \sum_{j=2}^{J} \binom{n}{j}^{-1} \sum_{1\le i_1<i_2<\cdots<i_j\le n} \tilde{\lambda}\{E(f_k; f_{i_1;k},\ldots,f_{i_j;k})\}$, where $E(f_k) =: E(f_k; f_{i_1;k},\ldots,f_{i_j;k}) = \{t \in I, \min_{r=i_1,\ldots,i_j} f_{r;k}(t) \le f_k(t) \le \max_{r=i_1,\ldots,i_j} f_{r;k}(t)\}$, $\tilde{\lambda}(f_k) = \lambda(E(f_k))/\lambda(I)$ and λ is the Lesbegue measure on I. The values of the modified band depth measure are stable with respect to the choice of J, and in order to be computationally faster we set $J=2$ and we denote $MBD_n^J(\mathbf{f})$ as $MBD(\mathbf{f})$ in the following. The use of the modified band depth measure avoids also having too many depth ties. Given the multivariate band depth measure defined in (2), a sample of multivariate functional data $\mathbf{f}_1,\ldots,\mathbf{f}_n$ can be ranked. In the following we denote $\mathbf{f}_{[i]}$ the sample curve associated with the ith largest depth value, so $\mathbf{f}_{[1]} = \mathrm{argmax}_{\mathbf{f}\in\{\mathbf{f}_1,\ldots,\mathbf{f}_n\}} MBD(\mathbf{f})$ is the *median* (deepest and more central) curve, and $\mathbf{f}_{[n]} = \mathrm{argmin}_{\mathbf{f}\in\{\mathbf{f}_1,\ldots,\mathbf{f}_n\}} MBD(\mathbf{f})$ is the most outlying one. Moreover, starting from the ranking provided for the sample of multivariate curves, multivariate functional boxplots can be obtained, in order to assess multivariate functional outliers.

3 Application to real and simulated ECG signals

In the following, we think to an ECG signal as a multivariate function $\mathbf{f}_i = (f_{i;1},\ldots,f_{i;8})$ with components $f_{i;k}$s ($k=1,\ldots,8$), generated by a stochastic process that takes values in $\mathbb{R}^8$. In other words it is considered as a signal composed by 8 correlated curves (leads I, II, V1, V2, V3, V4, V5 and V6) for each statistical unit (patient). We then apply the generalized concept of statistical depth to these multivariate functional data in order to rank them, according to the degree of deepness they present with respect to the entire sample. In [1, 2], an application of the tools briefly described above to a dataset of real ECG signals is proposed, considering 100 physiological traces, 50 Left Bundle Brunch Blocks (LBBBs) and 50 Right Bundle Brunch Blocks (RBBBs). These procedures can then be applied in order to carry out functional boxplots and to perform outliers detection for a new simulated ECG (physiological or pathological), whose deepness with respect to the corresponding population of signals we want to capture. Figure 1 shows the results on lead I of the multivariate functional boxplot for registered QT segments (further technical details are provided in [3]) of a sample of 100 phisiological and 1 simulated ECGs.

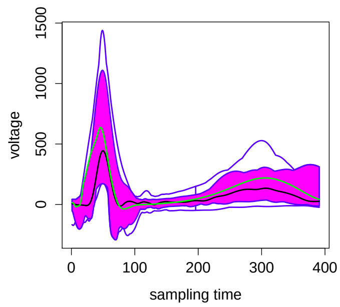

Fig. 1 Lead I of the multivariate functional boxplot for a simulated physiological ECG with respect to the corresponding population of the real physiological traces. Black solid line represents the functional median. Green solid line is the simulated ECG whose deepness we are interested in for validating numerical simulations.

We can then assess how much a simulated multivariate signal can be considered as coming from a reference population of signals, computing its depth in the way we described above. This is a new method for validating numerically simulated ECGs from a statistical perspective.

Acknowledgments The authors wish to thank Mortara Inc., 118 Dispatch Centre of Milan and the group of applied mathematics professor Jean Frédéric Gerbeau (Inria Paris-Rocquencourt, France) for having provided real and simulated data respectively.

References

1. Ieva, F., Paganoni, A.M.: Depth Measures for Multivariate Functional Data. *Submitted* [online] http://mox.polimi.it/it/progetti/pubblicazioni/quaderni/37-2011.pdf
2. Ieva, F.: Outlier detection for training sets in an unsupervised functional classification framework: an application to ECG signals. In: Proceedings of the 17th EYSM, Lisbon, Portugal, pp. 97–100 (2011)
3. Ieva, F., Paganoni, A.M., Pigoli, D., Vitelli, V.: Multivariate functional clustering for the analysis of ECG curves morphology. Tech. Rep. MOX, Math. Dept., Politecnico di Milano. (2011)
4. Li, J., and Liu, R.: New Nonparametric Tests of Multivariate Locations and Scales using Data Depth. Statistical Science **19**, 686–696 (2004)
5. Liu, R.: On a Notion of Data Depth Based on Random simplices. The Annals of Statistics, **18**, 405–414 (1990)
6. Lopez-Pintado, S., Romo, J.: On the Concept of Depth for Functional Data. Journal of the American Statistical Association, **104**, 486, 718-734 (2009)
7. Sun, Y., Genton, M.G.: Functional Boxplots, Journal of Computation and Graphical Statistics, to appear (2011)