MOX-Report No. 42/2021

Conformal Prediction Sets for Populations of Graphs

Calissano, A.; Fontana, M.; Zeni, G.; Vantini, S.

Conformal Prediction Sets for Populations of Graphs

Calissano[a], **Zeni**[b], **Fontana**[c], **Vantini**[d]

[a] MOX, Dipartimento di Matematica, Politecnico di Milano
Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
E-mail: anna.calissano@polimi.it
[b] MOX, Dipartimento di Matematica, Politecnico di Milano
Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
E-mail: gianluca.zeni@polimi.it
[c] MOX, Dipartimento di Matematica, Politecnico di Milano
Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
E-mail: matteo.fontana@polimi.it
[d] MOX, Dipartimento di Matematica, Politecnico di Milano
Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
E-mail: simone.vantini@polimi.it

Abstract

In the latest years, scholars started focusing on how to develop statistical tool for the analysis of population of complex data, such as sets of labelled or unlabelled graphs graphs. The present works adds to this literature by focusing on a strangely overlooked area, namely the formulation of prediction sets. By exploiting cutting edge techniques in the realm of machine learning, we propose a forecasting method for populations of both labelled and unlabelled graphs based on Conformal Prediction, able to identify prediction regions. Our method is model-free, achieves finite-sample validity, is computationally efficient and it identifies interpretable prediction sets, in the shape of a parallelotope. To explore the features of this novel forecasting technique, a simulation study and and a real-world example are presented.

Keywords: Population of Graphs; Conformal Prediction; Non-parametric Inference; Simultaneous Inference; Network Analysis

1. Introduction

One of the major challenges that the field of statistics is facing in the latest years, from both a theoretical and applied perspective, is to deal not only with a big and increasing amount of data, but also with statistical units of big and increasing complexity. As noted in Marron and Alonso (2014), the objects of the analysis for the modern statistician are no longer scalars or vectors, but new and complex

data structures or "objects", and this new field of research takes the name "Object-Oriented Data Analysis" (OODA). The first tasks that a statistician has to perform when dealing with this novel kind of statistical unit is to find a proper mathematical embedding and to "port" the standard elements of a statistician's toolbox (e.g clustering, principal component analysis, classification, regression). Starting from the pioneering analysis of functions embedded in an euclidean space (Ramsay, 1982; Ramsay and Silverman, 2005), statistical research has gradually shifted its focus towards statistical units of increasing complexity: from functions in non-euclidean spaces (Menafoglio et al., 2016), to trees (Billera et al., 2001; Wang and Marron, 2007; **?**) and graphs (Jain and Obermayer, 2009; Durante et al., 2017; Chowdhury and Mémoli, 2018; Calissano et al., 2020a).

Despite the relative wealth of methods present in the OODA literature, prediction and forecasting has been relatively overlooked, despite its importance in application tasks. While some attempts at tackling this problem have been proposed in a functional setting (Degras, 2011; Antoniadis et al., 2016) and for phylogenetic trees (Willis, 2019), to the best of our knowledge, no uncertainty quantification techniques has been proposed for more general graph data.

In this paper we are focusing on the analysis of populations of graphs (also referred as networks), namely a set of different graphs generated by the same data generating process. Hereby we focus on two classical scenarios of populations of graphs: populations of graphs either with the same nodes (labelled) or with different nodes (unlabelled). While the labelled graphs can be embedded in an euclidean space (labelled graphs are usually modelled as sets of adjacency matrices), unlabelled graphs require more complex geometrical embeddings such as Jain and Obermayer (2008); Chowdhury and Mémoli (2018). Notice that there can be many different scenarios between the pure labelled and the pure unlabelled, because many different matching between the nodes are possible. In this work we will focus on the Graph Space embedding proposed by Jain and Obermayer (2009), able to embrace both the pure labelled and the pure unlabelled graphs. Within this framework, some statistical models have been proposed such as Fréchet Means (Jain and Obermayer, 2008), Principal Components (Calissano et al., 2020a) and Linear Regression (**?**).

The aim of the present work is so to develop a set forecasting framework for populations of graphs. Due to the natural complexity of a population of graphs, we adopt a Conformal Prediction (Vovk et al., 2005; Zeni et al., 2020) framework. Apart from the evident theoretical interest, forecasting a new instance of a population of graphs has also a clear and meaningful application in different fields, such as in the analysis of international trade networks (Amador and Cabral, 2017), Input-Output networks (Cerina et al., 2015) and epidemic models where the subjects are connected using a network topology (Ball et al., 2019).

In Section 2, we will introduce population of labelled graphs. Section 3 is then devoted to the definition of the concept and desirable shape of a Conformal Prediction set in the population of labelled graphs case. To tackle an issue of having sets of constant amplitude for all the edges and nodes, in Section 3.2 we modify the procedure via the use of a modulation function, which lets to adjust the amplitude

of the identified set in a data-driven fashion. Section 4 generalizes the presented framework to the much more challenging case of population of unlabelled graphs. We then present a simulation study, in which we explore the performance of our method (Section 5.1), by analysing labelled and unlabelled graphs and comparing different parametric and non-parametric prediction intervals and. An application to a real world scenario is provided in Section 5.4, where we apply the developed tools to a dataset describing human mobility during COVID-19 in Lombardy.

2. Mathematical Structure of Populations of Graphs

Let $X = (V, E, a)$ be a graph (or a network) with vertex set $V = \{v_1, \ldots, v_n\}$. The edge set is $E \in V \times V$, and its size is $n^2 = p$. $a : E \to A$ represents the attribute map. All the following theoretical framework is presented for graphs with scalar euclidean attributes on nodes and edges ($A = \mathbb{R}$). Some notions about the possible generalizations of this assumption are presented in Section 6. Notice that the node attributes are described as self-loops. I this framework, graphs are represented as flattened adjacency matrices $x \in X$, $X = \mathbb{R}^p$, where $p = n^2$. Given a set of graphs $x_1, \ldots, x_k, x_i \in X$, the distance between two flattened adjacency matrices is defined as the sum of the distances between the attributes on nodes and edges:

$$d_X(x_1, x_2) = \sum_{j=1}^{p} d(x_1(j), x_2(j)) \tag{1}$$

where $d : A \times A \to \mathbb{R}$ is the chosen distance between the attributes, and $x(j)$ is the $j-th$ element of the vectorized graph $x \in X$. In our framework, d will be a squared euclidean distance.

If the nodes have a certain unique and non interchangeable meaning across the graphs, the graphs are said to be *labelled*. If, instead, the nodes are interchangeable and their meaning is not unique but it is related to the role they assume within the graph topology, we have *unlabelled* graphs. An example of labelled graphs is a set of graphs measuring the social interaction between the same individuals along time. An example of unlabelled graphs is a set of social interaction between students of the same age in different classes (i.e. one graph for each class). The reader should note how a graph being labelled or unlabelled is a problem driven decision and depends on the aims and scope of the analysis. For example, the graph of social interaction between the same people along time can be interpreted as unlabelled if the researcher is interested in finding interchangeable role between individuals along time, and as labelled if the same friends are present in all the social networks.

3. Prediction Parallelotopes for Labelled Graphs

Consider an i.i.d. population of graphs $X_1, \ldots, X_k, x_i \in X$, sampled from a distribution $\mathbb{P}$. The problem we want to tackle is define an interval for an estimator.

Formally, we define a prediction set $\mathcal{C}_{k,1-\alpha} := \mathcal{C}_{k,1-\alpha}(X_1, \ldots, X_k)$ as

$$\mathbb{P}(X_{k+1} \in \mathcal{C}_{k,1-\alpha}) \geq 1 - \alpha \tag{2}$$

where $\alpha \in [0, 1]$. The measure on X is the Lebesgue measure defined on the Borel σ−algebra. Let $X(j), j = 1, \ldots, p$ be a generic element of the flattened adjacency matrix X, where $p = n^2$ since the attributes are scalars, and we are working with weighted adjacency matrices. As we said previously, $X(j) \in \mathbb{R}$. With respect to a univariate setting, the case of formulating prediction sets for complex data poses a serious question in terms of interpretability and practical usefulness of the obtained intervals. It is intuitive to understand that the best case in terms of interpretability for a prediction set is a region in space that allows a component-wise identification of an element that is inside or outside the prediction region. In more mathematical terms, we are interested in a set defined as:

$$\mathcal{C} := \{X \in X : X(j) \in \mathcal{C}(j), \quad \forall j \in 1, \ldots, p\}, \tag{3}$$

where $\mathcal{C}(j) \subseteq \mathbb{R}$. The sets described in Equation 3 are the Cartesian product of p intervals of the real line. A set like this forms a parallelotope in $\mathbb{R}^p$

The prediction set in the shape of a parallelotope allows a practitioner to project the multivariate prediction region which is valid at a level α, in intervals for each element of X without changing the coverage level. Our applied goal is thus to identify parallelotope-shaped sets with a given unconditional coverage level, namely: $\mathbb{P}(X_{k+1} \in \mathcal{C}_{k,1-\alpha}) \geq 1 - \alpha$.

3.1 Conformal Prediction Parallelotopes for Graphs

A method which has the explicit aim to identify prediction sets of the type described in Equation 3 is the Conformal Prediction Method (Vovk et al., 2005; Zeni et al., 2020). The key quantity around which a Conformal Prediction framework revolves is a so called conformity (or non-conformity) measure, on which a very weak "quasi-model" (in the sense of Cella and Martin, 2020) is imposed, and it allows to obtain prediction sets. By aptly choosing a non-conformity measure with the desired iso-contours, one is able to obtain prediction sets with the minimal size and/or with the desired shape.

Let us put ourselves in a Split/Inductive Conformal Framework (see Papadopoulos et al. (2002); Lei and Wasserman (2014) for an introduction). We start by splitting our $X_1, \ldots, X_k$ in a proper *training set* $\mathcal{I}_1$ and a *calibration set* $\mathcal{I}_2$, where $|\mathcal{I}_1| + |\mathcal{I}_2| = k$, indexing the sets as: $l \in \mathcal{I}_1$ and $m \in \mathcal{I}_2$.

$\forall X_m, m \in \mathcal{I}_2$, one can compute an empirical P-value:

$$p_{X_m} := \frac{|\{i \in \mathcal{I}_2 : R_i \geq R_m\}|}{|\mathcal{I}_2| + 1}$$

where $R : X^k \to \mathbb{R}$ is a non-conformity measure as defined in Vovk et al. (2005). The conformal prediction set defined using the above definition of P-value is:

$$\mathcal{C}_{k,1-\alpha} := \{X \in X : p_X > \alpha\} \tag{4}$$

4

where α is the desired coverage level. To identify a prediction set that is also a set in the sense of Equation 3, we can use the $L_{\inf}$ metric. Namely, we define our non-conformity measure (NCM) $R : X^k \rightarrow \mathbb{R}$ to be

$$R_m = \max_{j \in 1,\ldots,p} \left| X_m(j) - \hat{\mu}(j) \right|, \quad m \in \mathcal{I}_2 \tag{5}$$

where $\hat{\mu} = \mathcal{A}(\{X_l, l \in \mathcal{I}_1\})$ is an estimator of central tendency based on a algorithm $\mathcal{A}$, trained on $X_l, l \in \mathcal{I}_1$. We can also note that, having defined a prediction set as in Equation 10, and the NCM as in Equation 5, one can say that $X_{k+1} \in \mathcal{C}_{k,1-\alpha} \iff R_{k+1} \leq h$, with h the $\lceil (|\mathcal{I}_2| + 1)(1 - \alpha) \rceil$-th smallest value in the set $\{R_m : m \in \mathcal{I}_2\}$. Then

$$\max_{j \in 1,\ldots,p} \left| X_{k+1}(j) - \hat{\mu}(j) \right| \leq h$$
$$\Rightarrow \left| X_{k+1}(j) - \hat{\mu}(j) \right| \leq h \quad \forall(j)$$
$$\Rightarrow X_{k+1}(j) \in [\hat{\mu}(j) - h, \hat{\mu}(j) + h] \quad \forall j = 1,\ldots,p$$

Therefore, the split conformal prediction set induced by the nonconformity measure (5) is

$$\mathcal{C}_{k,1-\alpha} := \{X \in X : X(j) \in [\hat{\mu}(j) - h, \hat{\mu}(j) + h] \quad \forall j = 1,\ldots,p\}. \tag{6}$$

Algorithm 1 Split Conformal Prediction Parallelotopes for Populations of Graphs

---

1: **Require:** Data $X_i, i = 1,\ldots,k,$, type-1 error level $\alpha \in (0,1)$, central tendency estimation algorithm $\mathcal{A}$
2: split randomly $\{1,\ldots,k\}$ into two subsets $\mathcal{I}_1, \mathcal{I}_2$
3: $\hat{\mu} = \mathcal{A}(\{X_l, l \in \mathcal{I}_1\})$
4: $R_m = \max_{j=1,\ldots,p} (X_m(j) - \hat{\mu}(j)), m \in \mathcal{I}_2$
5: h is the $\lceil (|\mathcal{I}_2| + 1)(1 - \alpha) \rceil$-th smallest value in the set $\{R_m : m \in \mathcal{I}_2\}$
6: **Ensure:** $\mathcal{C}_{k,1-\alpha} := \{X \in X : X(j) \in [\hat{\mu}(j) - h, \hat{\mu}(j) + h] \quad \forall j\}$

The described procedure is summarised in Algorithm 1. The calculation of these sets is very convenient: we require to train the central tendency estimation algorithm $\mathcal{A}$ only once, and we have a closed form for the calculation of the semi-amplitude of the set.

3.2 Amplitude Modulation

The main shortcoming of the approach proposed in Section 3.1 is that the identified parallelotope has constant amplitude across all $j = 1,\ldots,p$. While there may be situations in which such feature is desirable, practitioners usually face cases in which edge attributes have different variability, and may want to take into account such variability when making a global prediction (for instance, with wider or narrower sets). The limit situation in this case would be the one where a vertex is

5

completely absent from the population of graph analysed, meaning that $X(j) = 0$ for the position j assigned to the null node. Any amplitude different from 0 for the interval in $B(j)$ is not desirable, since we do not want to "add" a node that never appears in the observed data.

To do so, following Lei et al. (2018), we condition the amplitude of Equation 6 across $j \in 1, \ldots, p$ using a local notion of variability. Namely, we modify Equation 5 in the following fashion:

$$R_m = \max_{j \in 1, \ldots, p} \left[\frac{\left| X_m(j) - \hat{\mu}(j) \right|}{\hat{s}(j)} \right] \tag{7}$$

Where $\hat{s} = \mathcal{S}(\{X_l, l \in \mathcal{I}_1\})$ is an estimator of local variability, trained on the set $\{X_l, l \in \mathcal{I}_1\}$ using the symmetric algorithm $\mathcal{S}$. Please note that the modulation function is computed on the training set only.

We summarise the Split Conformal Procedure with Modulation in Algorithm 2

Algorithm 2 Split Conformal Prediction Parallelotopes for Populations of Graphs with Amplitude Modulation

---

1: **Require:** Data X_i, $i = 1, \ldots, k$,, type-1 error level $\alpha \in (0, 1)$, regression algorithm $\mathcal{A}$, amplitude modulation algorithm $\mathcal{S}$
2: split randomly $\{1, \ldots, k\}$ into two subsets $\mathcal{I}_1, \mathcal{I}_2$
3: $\hat{\mu} = \mathcal{A}(\{X_l, l \in \mathcal{I}_1\})$, $\hat{s} = \mathcal{S}(\{X_l, l \in \mathcal{I}_1\})$
4: $R_m = \max_j \left(\frac{|X_m(j) - \hat{\mu}(j)|}{\hat{s}(j)} \right)$, $m \in \mathcal{I}_2$
5: $h =$ is the $\lceil (|\mathcal{I}_2| + 1)(1 - \alpha) \rceil$-th smallest value in the set $\{R_m : m \in \mathcal{I}_2\}$
6: **Ensure:** $\mathcal{C}_{k, 1-\alpha} := \{X \in X : X(j) \in [\hat{\mu}(j) - h\hat{s}(j), \hat{\mu}(j) + h\hat{s}(j)] \quad \forall j = 1, \ldots, p\}$

According to the choice of the algorithm to compute $\mathcal{S}$, the modulating behaviour dramatically changes. We mention two notable cases: (i) When $\hat{s}(j) = 1 \, \forall j$ and no modulation is taking place, Algorithm 2 will yield results equal to Algorithm 1; (ii) If $\hat{s}(j) = \sqrt{Var(X(j))}$ the resulting set amplitude will be modulated according to the local variability of the attributes of the graph.

Remark 1 *Consider the case when $\hat{s}(j) = \sqrt{Var(X(j))}$, if an attribute of an element of the graph is deterministic, the attentive reader can immediately understand that both the numerator and the denominator of the non-conformity function will be equal to zero. This yields to an indeterminate form. To solve this computational problem, we will use, as a modulation function $\hat{s}(j) = \sqrt{Var(X(j))} + \epsilon$, where ϵ is a very small constant. This yields to a prediction interval for a deterministic values centered in the actual value of the constant, and of negligible length, while allowing the computation of the global set to be performed.*

4. Prediction Class of Parallelotopes for Unlabelled Graphs

In this section, we are going to extend the concept introduced in the previous section to the more general, unlabelled, framework. As already stated in Section 1, a

set of graphs can share the same nodes, share only a subset of nodes, or not share any nodes. If the nodes are not the same, graphs are considered as *unlabelled* and they should be compared using some kind of isomorphism, allowing a meaningful topological comparison. In this paper, the specific isomorphism is obtained by applying a permutation action. This framework has been introduced in a general form by Jain and Obermayer (2009) and studied in the particular case of graphs by Guo et al. (2019); Calissano et al. (2020a). Note that the choice of permutation actions to compare unlabelled graphs is not necessarily the only possible one. There are in fact other group action used to compare graph structures such as the one introduced by Chowdhury and Mémoli (2017).

Applying the nodes permutation group action T to the space of flattened adjacency matrices X gives rise to a quotient space X/T, called Graph Space (Jain and Obermayer, 2009; Calissano et al., 2020a), where every equivalence class corresponds to all the graphs you can obtain from one adjacency matrix by permuting the nodes. This geometrical framework has the labelled setting as the special case that arises when the only allowed permutation is the identity.

From the definition of the distance in the X space, we can define the following distance in X/T:

$$d_{X/T}(X_1, X_2) = \min_{t \in T} d_X(tX_1, X_2) \tag{8}$$

Where $t \in T$ is a permutation matrix of $\{0, 1\}$ values applied to the flatted adjacency matrix. Given a $t \in T$, we can associate a unique function $\sigma_t : V \rightarrow V$ that assign the corresponding permuted index to the original one. All the geometrical properties of X/T are thoroughly described in Calissano et al. (2020a). We remind the reader that the Graph Space is a geodesic metric measure space, but it is not a manifold. Here we report only the needed definitions for the purpose of this article. In particular, it is useful to define the projection function $\pi : X \rightarrow X/T$ that associates every element to its equivalent class $x \mapsto [x] = \{tx, \forall t \in T\}$. We also remind the reader the push-forward measure on X/T given a measure on X:

Definition 2 *The Graph Space X/T is endowed with a probability measure η which is absolutely continuous with respect to the push-forward of the Lebesgue measure m on X. In particular, for $A \subset X/T$, we have $\eta(A) = 0$ if $m(\pi^{-1}(A)) = 0$.*

Definition 3 *A set $C \in X$ as defined in Equation 3 can be projected on the Graph Space as:*

$$[\mathcal{C}] = \bigcup_{t=1}^{|T|} \bigtimes_{j=1}^{p} \mathcal{C}(\sigma_t(j)), \quad [\mathcal{C}] \subseteq X/T$$

where $\sigma_t : \{1, \ldots, p\} \rightarrow \{1, \ldots, p\}$ is the relabelling function associated to the permutation $t \in T$

The idea is to define a set of intervals that follow the index permutation of the elements in the Graph Space. For the sake of simplicity we define:

$$\mathcal{C}^t = \bigtimes_{j=1}^{p} \mathcal{C}(\sigma_t(j))$$

7

where $\mathcal{C}(\sigma_t(j)) \subseteq \mathbb{R}$, $\mathcal{C}^t \subseteq \mathbb{R}^p$.

The probability of this interval in the Graph Space can be computed using the projection on the total space X:

$$\mathbb{P}_\eta \left(\bigcup_{t=1}^{|T|} \mathcal{C}^t \right) = \sum_{t=1}^{|T|} \left((-1)^{t-1} \sum_{\substack{I \subseteq \{1,\dots,|T|\} \\ |I|=t}} \mathbb{P}(A_I) \right)$$
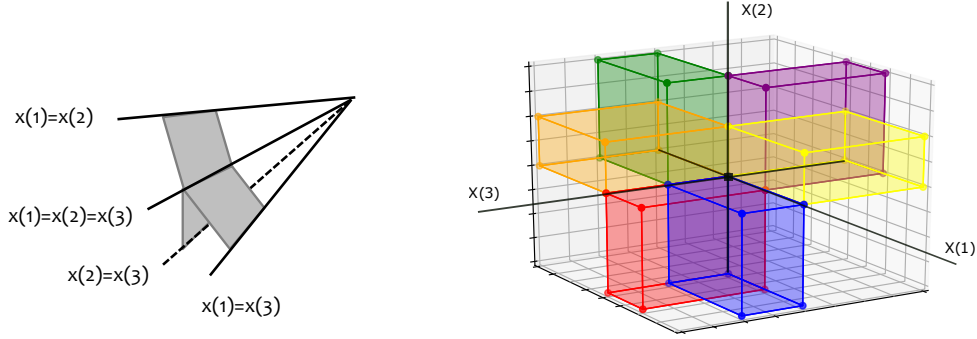
where $A_I := \bigcap_{t \in I} \mathcal{C}^t$

Figure 1: Conceptual visualization of the shape of an interval in the Graph Space and its back-projection π^{-1} in the total space for an un-directed graph with three nodes and no attributes on the nodes.

Example 1 *Consider an undirected graph with three nodes $n = 3$, real attributes on edges and no attributes on nodes. This graph can be described as a point in $\mathbf{R}^3$. The number of permutation is $3! = 6$. The interval $\mathcal{C}$ is the Cartesian product of three intervals on the real line $\mathcal{C} = \mathcal{C}(1) \times \mathcal{C}(2) \times \mathcal{C}(3)$. If we permute this shape with the following permutation $t = \{2,1,3\}$, we obtain a new set $\mathcal{C}^t = \mathcal{C}(\sigma_t(1) = 2) \times \mathcal{C}(\sigma_t(2) = 1) \times \mathcal{C}(\sigma_t(3) = 3)$. The union of all their possible permutations is $[\mathcal{C}]$, shown in Figure 1*

As in the labelled case, we can define the interval with a given coverage level:

$$\mathbb{P}_\eta \left([X]_{k+1} \in [\mathcal{C}]_{k,1-\alpha} \right) \geq 1 - \alpha \tag{9}$$

We start by splitting our set of unlabelled graphs $\{[X_1], \dots, [X_k]\}$ in a training set $\mathcal{I}_1$ and a calibration set $\mathcal{I}_2$, where $|\mathcal{I}_1| + |\mathcal{I}_2| = k$.

$\forall [X_m], m \in \mathcal{I}_2$, one can compute an empirical P-value define exactly as in the labelled case:

$$p_{[X_m]} := \frac{|\{i \in \mathcal{I}_2 : R_i \geq R_m\}|}{|\mathcal{I}_2| + 1}$$

where R is a non-conformity measure as defined by Vovk et al. (2005). The conformal prediction set defined using the above definition of P-value can be identified as

$$[\mathcal{C}]_{k,1-\alpha} := \left\{ [X] \in X/T : p_{[X]} > \alpha \right\} \tag{10}$$

8

We define our non-conformity measure R_m to be

$$R_m = \max_{j=1,\dots,p} |t_m X_m(j) - \mathcal{A}(j)| \tag{11}$$

where:

$$t_m = \arg\min_{t \in T}(d_X(tX_m, \mathcal{A})) \tag{12}$$

where $\mathcal{A}$ is a symmetric function of the data. This non-conformity measure selects the permutation that optimally aligns the two graphs and consequently selects the edge or node that are mostly far apart from each other.

In the case of amplitude modulation, the Equation 11 becomes:

$$R_m = \max_{j=1,\dots,p} \frac{|t_m X_m(j) - \mathcal{A}(j)|}{\hat{s}(j)} \tag{13}$$

Where $\hat{s}$ is an estimator of the variability of the edge or node j after the alignment with respect to the central estimator $\mathcal{A}$. The whole procedure is summarised in Algorithm 3.

Algorithm 3 Split Conformal Prediction Parallelotopes for Populations of Unlabelled Graphs with Amplitude Modulation

---

1: **Require:** Data $[X_i]$, $i = 1,\dots,k,$, type-1 error level $\alpha \in (0,1)$, Predictive algorithm $\mathcal{A}$, amplitude modulation algorithm s
2: split randomly $\{1,\dots,k\}$ into two subsets $\mathcal{I}_1, \mathcal{I}_2$
3: $\hat{\mu} = \mathcal{A}(\{[X_l], l \in \mathcal{I}_1\})$, $\hat{s} = \mathcal{S}(\{[X_l], l \in \mathcal{I}_1\})$
4: Find $\{t_1, \dots, t_{|\mathcal{I}_2|}\}$ s.t. $t_m = \arg\min_{t \in T}(d_X(t_m X_m, \hat{\mu}))$
5: $R_m = \max_{j=1,\dots,p}\left(\frac{|(t_m X_m)(j) - \hat{\mu}(j)|}{\hat{s}(j)}\right) = \max_{j=1,\dots,p}\left(\frac{|X_m(\sigma_{t_m}(j)) - \hat{\mu}(j)|}{\hat{s}(j)}\right)$, $m \in \mathcal{I}_2$
6: h is equal to $\lceil(|\mathcal{I}_2| + 1)(1 - \alpha)\rceil$-th smallest value in the set $\{R_m : m \in \mathcal{I}_2\}$
7: **Ensure:** $\mathcal{C}_{k,1-\alpha} := \{[X] \in X/T : (tX)(j) \in [\hat{\mu}(j) - h\hat{s}(j), \hat{\mu}(j) + h\hat{s}(j)] \quad \forall j = 1,\dots,p, \quad \forall t \in T\}$

Remark 4 *The reader should note how it is not required to specify anything about $\mathcal{A}$. This generality, which is one of the main interesting features of the conformal prediction framework, allows for the use of any predictive algorithm, either statistical inspired, machine-learning inspired, or a combination of the two. However, due to the geometrical complexity of the Graph Space, extending regression strategies as well as neural network strategy to this framework is not straightforward. In this paper, we are going to use the Fréchet Mean as the $\mathcal{A}$ (see (Calissano et al., 2020a) for definitions and details).*

5. Simulation and Case Study

In this section, we illustrate the theoretical results described in the previous section on two simulated dataset and one case studies. In all these examples, the $\mathcal{A}$ function is going to be a Frèchet Mean estimator. In the unlabelled case, the

Frèchet mean corresponds to the Sample Mean, while in the unlabelled case, the Fréchet Mean is computed with the Align All and Compute Procedure (Calissano et al., 2020a). The Conformal Prediction Parallelotopes is implemented as a function in the Python Package *Graph Space* (Calissano et al., 2020b).

5.1 Simulation

5.2 Simulation: Labelled Case

In this simulation, we compute the Empirical Coverage of different parametric intervals and the conformal prediction intervals. We generated a set of graphs 130 graphs ($|\mathcal{I}_1| + |\mathcal{I}_2| = 30$ is the training set - eventually divided in training and calibration for the split conformal method- and 100 is the test set) for 100 times. Every directed graph has 5 nodes with Gaussian attributes $N(0,1)$ and 20 edges following four different distributions:

1. Gaussian attributes $N(0,1)$,

2. Uniform attributes $U(-1.7, 1.7)$,

3. t-Student attributes with 4 degrees of freedom,

4. t-Student attributes with 1 degree of freedom,

Having two different distributions on nodes and edges attributes is very common in the applications, because nodes and edges usually describe two different phenomena. For every generated model, we compute the sample mean and three different prediction intervals, in a labelled fashion:

1. Univariate Gaussian Intervals:

$$\hat{x}_i \pm t_{k-1}(\alpha/2)\sqrt{1 + \frac{1}{k}}\hat{s}_i$$

2. Univariate Gaussian Intervals with Bonferroni Correction:

$$\hat{x}_i \pm t_{k-1}((\alpha/p)/2)\sqrt{1 + \frac{1}{k}}\hat{s}_i$$

3. Simultaneous Gaussian Intervals:

$$\bar{x}_i \pm \hat{s}_i\sqrt{\left(1 + \frac{1}{k}\right)\frac{(k-1)p}{(k-p)}F_{(p,k-p)}(\alpha)}$$

4. Gaussian Ellipse

$$(x - \bar{x})'\hat{S}^{-1}(x - \bar{x}) \le \left(1 + \frac{1}{k}\right)\frac{(k-1)p}{(k-p)}F_{(p,k-p)}(\alpha)$$

where $t_{k-1}(\alpha/2)$ and $F_{(p,k-p)}(\alpha)$ denotes the upper quantile. $\hat{S}$ denotes the sample covariance matrix and $\hat{s}_i$ the estimated sample variance; $\hat{x} = [\hat{x}_1, \ldots, \hat{x}_p]$ the sample vector mean.

Given a set of 100 different different α, we compute the empirical coverage on the test set, defined as:

$$\hat{E}(1-\alpha) = \frac{1}{100} \sum_{k=1}^{K} \sum_{i=0}^{100} \frac{1_{x_i \in IC_k(\alpha)}}{100} \tag{14}$$

In the Figures 2, we show the calibration curves α for the the different generative models and the different intervals. As expected, and coherently with the
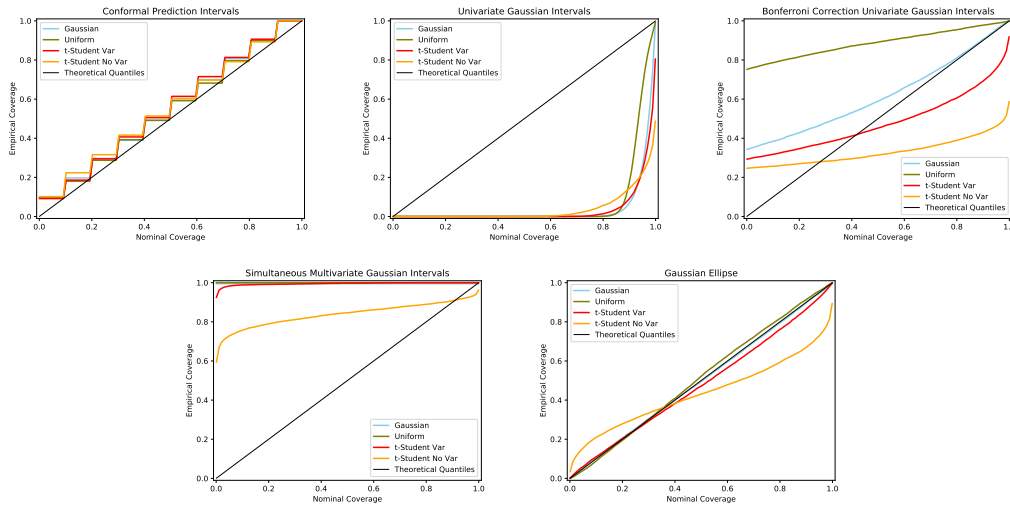
Figure 2: Empirical Coverage as a function of theoretical α

theory, Univariate gaussian intervals are wildly under-covering in a global sense, given the absence of any multiplicity correction. Bonferroni-corrected intervals are quite conservative in the gaussian case, and very conservative in leptocurtic cases such as the uniform one. With respect to the two platicurtic cases (t-student with 1 degree of freedom and t-student with 4 degrees of freedom), we observe a generic conservativeness for low levels of nominal coverage, generated by the Bonferroni correction: this effects tends to disappear for higher coverage levels. We see that Bonferroni-corrected intervals are under-covering for nominal levels that are commonly used in the practice. The projection over the components of multivariate gaussian intervals generates, similarly to our method, prediction sets with the shape of a parallelotope: It appears evident how they are grossly conservative. Multivariate gaussian prediction ellipses are exact in the gaussian case, and provide a conservative approximation in the uniform case, while they fail to cover for platicurtic distributions. In any case a high-dimensional prediction ellipse such as this one is of very relative practical value. The only method that is able to provide properly calibrated regardless of the distribution and interpretable (thanks to their parallelotopic shape) prediction intervals is the proposed conformal one.

11

5.3 Simulation: Unlabelled Case

In this example, we simulated 500 pentagons from the equivance classes shown in Figure 5b. We first randomly select one of the equivalent class and then we randomly pick one element from the class (i.e. a random permutation of the graphs shown in the figure). These pentagons have constant attribute on nodes (10) and decreasing attributes on edges (100, 80, 60, 40, 20). The interesting part of this example reside on the simplicity of the graph topology and the possibility of visualizing and understanding the correct alignment as shown in Figure 5b.
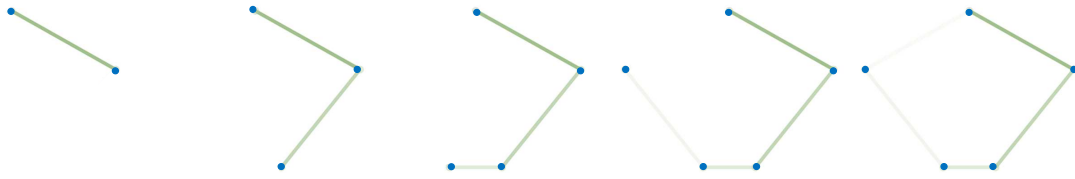
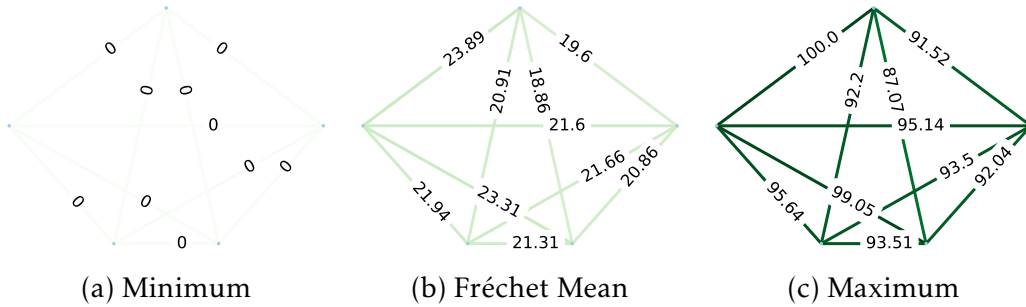

Figure 3: Example of Unlabelled Dataset

(a) Minimum   (b) Fréchet Mean   (c) Maximum

Figure 4: Labelled Setting: Conformal Prediction interval of level 95%.

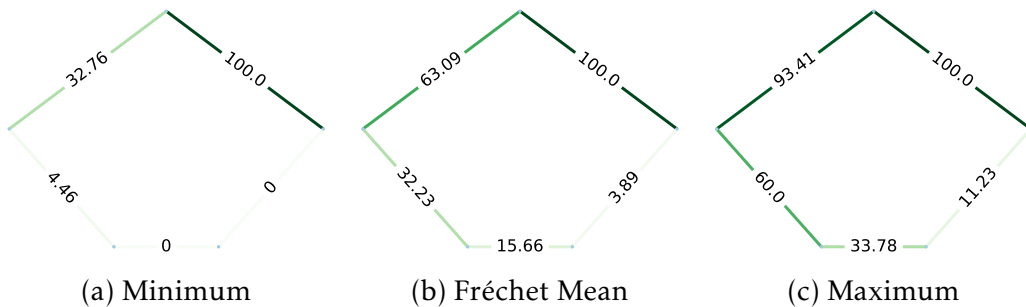(a) Minimum   (b) Fréchet Mean   (c) Maximum

Figure 5: UnLabelled Setting Conformal Prediction interval of level 95%.

In Figures 4 and 5, we show the Fréchet Mean and the corresponding 95% intervals for both the labelled (no alignment procedure is applied to the data) and

the unlabelled setting. The exact values of the intervals are visualized on top of each edge. The example shows the capability of the conformal prediction intervals of capturing the topology and the attribute, when working in an unlabelled setting.

5.4 Application

In this section, we apply the described methodology to case study concerning the mobility during the first phases of the COVID-19 epidemic in the Italian region of Lombardy. Covid-19 hit with particular violence the northern part of Italy. The Italian government decided for a complete lockdown from the 21st of March 2020. The so-called phase II started the 4th of May 2020, with a slow reopening of commercial activities. The data are provided by a location marketing company Cuebiq and consists of GPS location data gathered via smartphone of anonymous users in the Lombardy Region from the 17th of February to the 17th of May 2020. Among the anonymous users we randomly sampled 50000. This anonymised data is collected from users who opt-in to share their data for research purposes, through a GDPR-compliant framework. Cuebiq then applies additional privacy preservation techniques to remove sensitive locations from the dataset, and to obfuscate personal areas such as home locations by "up-leveling" them to 600m x 600m geo-hash tiles. Data have been aggregated in Origin-Destination Matrices (ODM). ODMs - a standard data type commonly used in transport and mobility modelling - are graphs where nodes are geographical locations and edges are the flows of people between locations. In this case study, we focus on the peoples' trips arriving before 7 p.m. of the working days of the given period. The result is a datset of 65 labelled graphs with 11 nodes each. In Figure 6, the map of the 11 provinces of Lombardy and an example of a ODM is reported.

The conformal prediction intervals for the Fréchet Mean are computed at three different levels of $\alpha = 0.25, 0.5, 0.75$, to shows how this conformal prediction strategy can be used to understand the distribution of a complex phenomena. In Figure 7, we report the intervals of the incoming and outgoing edges from the province of Bergamo (cut for negative values and standardized due to privacy reasons). Bergamo was the province most hit by the COVID-19 epidemic in Lombardy and it is an important regional and national economic hub, being at the centre of a very industrialized area. As it is clear from the prediction interval, the COVID lockdown decreased the mobility to and from Bergamo province and the mobility activities have not recovered in the *Phase II*. From a modelling perspective, the plots shows how the intervals sizes increase for higher values of $1 - \alpha$. Notice that the prediction intervals are in dimension 144 and in the Figure 7 we are only showing the components along 11 axes, even if the coverage is in the higher dimension.

6. Discussion and Conclusions

The issue of predicting with uncertainty the complex statistical units analysed in Object Oriented Data Analysis is a key research topic in modern statistics, from

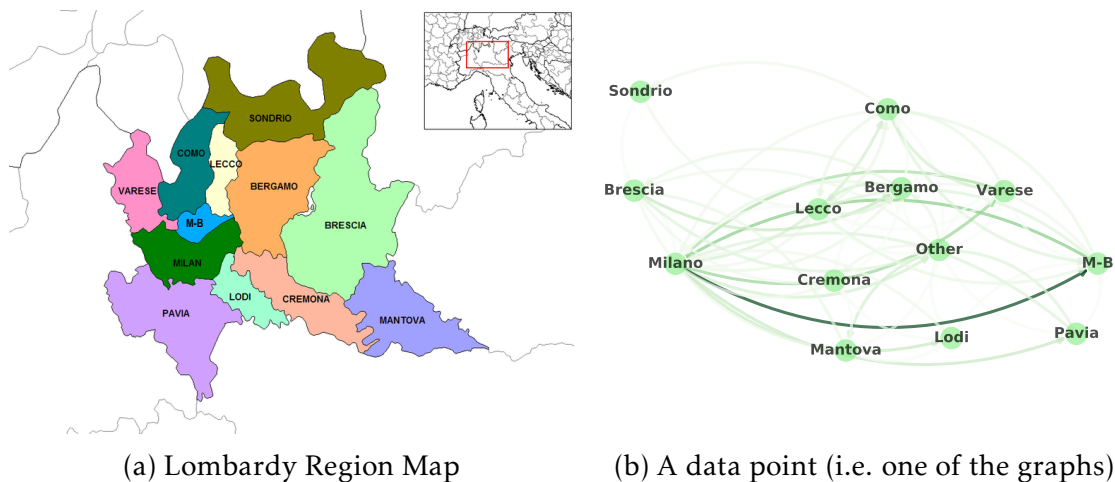(a) Lombardy Region Map　　　　(b) A data point (i.e. one of the graphs)

Figure 6: A map of the Lombardy Region with the 11 provinces (M-B stands for Monza-Brianza) and an example of data point randomly sampled from the Origin-Destination Population of graphs.
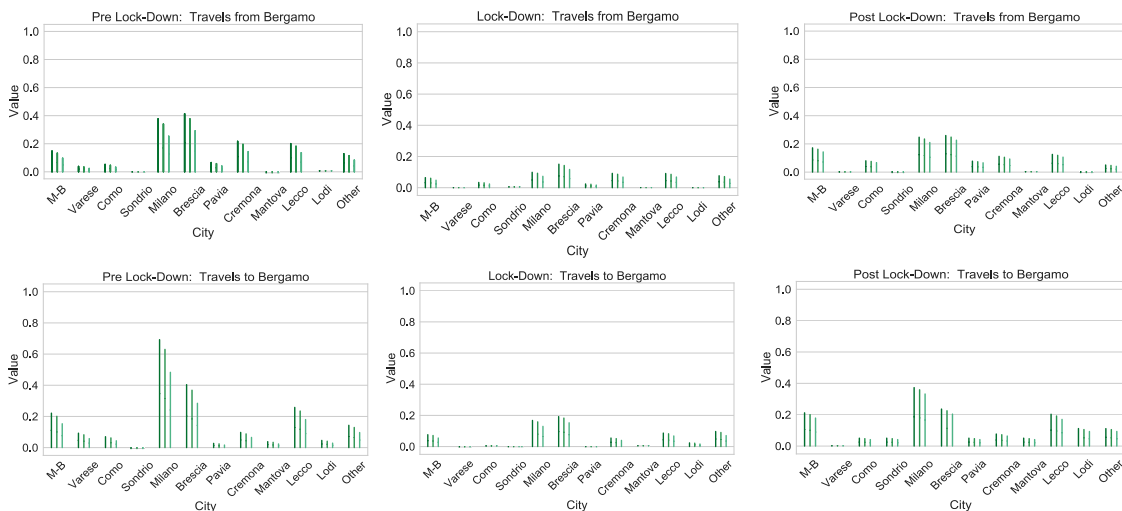

Figure 7: Intervals with $\alpha = 0.25, 0.5, 0.75$ for the outgoing and incoming edges of the province of Bergamo. Data has been standardized and negative values are set to zero.

both a theoretical and applied perspective. Only a few works have been proposed on the topic, which focus on set forecasting techniques based on either distributional assumptions that are hard to justify, or on heavy computational methods. Moreover, they concentrate on producing forecasts for well-behaved objects, for which an embedding in a euclidean or mildly non-euclidean space is possible.

In this work, we address this series of problems by proposing a model-free, computationally efficient set forecasting method, based on Conformal Prediction,

14

for two types of statistical objects of increasing complexity: population of labelled and unlabelled graphs with scalar attributes on nodes and edges. It should be noted that the current framework can be extended to population of graphs with more complex euclidean (and potentially even non euclidean) attributes.

An extensions of the current work regards the embedding strategy, and in particular the codification of absent nodes and edges as zeros. This choice is very convenient because it allows an euclidean representation of the same dimension for the adjacency matrices but it treats equally absent elements with existing elements with zero attributes. The problem is not relevant in the context where a zero attributes correspond to a non-existing element (e.g. context were the weight represent the intensity of the relation). Otherwise, conformal prediction can be extended to other embedding context such as the one proposed in Chowdhury and Mémoli (2018).

The current work poses the basis of the conformal prediction for population of graphs. As soon as novel statistical prediction methods will be defined to the context of population of graphs, the current framework can be easily extended.

Acknowledgements

References

J. Amador and S. Cabral. Networks of Value-added Trade. *The World Economy*, 40(7): 1291–1313, 2017.

A. Antoniadis, X. Brossat, J. Cugliari, and J.-M. Poggi. A prediction interval for a function-valued forecast model: Application to load forecasting. *International Journal of Forecasting*, 32(3):939–947, July 2016.

F. Ball, T. Britton, K. Y. Leung, and D. Sirl. A stochastic SIR network epidemic model with preventive dropping of edges. *Journal of Mathematical Biology*, 78(6):1875–1951, 2019.

L. J. Billera, S. P. Holmes, and K. Vogtmann. Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics*, 27(4):733–767, 2001.

A. Calissano, A. Feragen, and S. Vantini. Populations of unlabeled networks: Graph space geometry and geodesic principal components. *MOX Report*, 2020a.

A. Calissano, A. Feragen, and S. Vantini. Graphspace python package. *https://github.com/annacalissano/GraphSpace.git*, 2020b.

L. Cella and R. Martin. Valid distribution-free inferential models for prediction. *arXiv:2001.09225*, 2020. arXiv: 2001.09225.

F. Cerina, Z. Zhu, A. Chessa, and M. Riccaboni. World Input-Output Network. *PLOS ONE*, 10(7):e0134025, 2015. Publisher: Public Library of Science.

S. Chowdhury and F. Mémoli. Distances and isomorphism between networks and the stability of network invariants. *arXiv preprint arXiv:1708.04727*, 2017.

S. Chowdhury and F. Mémoli. The metric space of networks. *arXiv preprint arXiv:1804.02820*, 2018.

D. A. Degras. Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, 21(4), 2011.

D. Durante, D. B. Dunson, and J. T. Vogelstein. Nonparametric Bayes modeling of populations of networks. *Journal of the American Statistical Association*, 112(520):1516–1530, 2017.

X. Guo, A. Srivastava, and S. Sarkar. A quotient space formulation for statistical analysis of graphical data. *arXiv preprint arXiv:1909.12907*, 2019.

B. Jain and K. Obermayer. On the sample mean of graphs. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 993–1000. IEEE, 2008.

B. J. Jain and K. Obermayer. Structure spaces. *Journal of Machine Learning Research*, 10: 2667–2714, 2009.

J. Lei and L. Wasserman. Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):71–96, 2014.

J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman. Distribution-Free Predictive Inference for Regression. *Journal of the American Statistical Association*, 113(523): 1094–1111, 2018.

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

A. Menafoglio, O. Grujic, and J. Caers. Universal Kriging of functional data: Trace-variography vs cross-variography? application to gas forecasting in unconventional shales. *Spatial Statistics*, 15:39–55, 2016.

H. Papadopoulos, K. Proedrou, V. Vovk, and A. Gammerman. Inductive Confidence Machines for Regression. In *Machine Learning: ECML 2002*, Lecture Notes in Computer Science, pages 345–356. Springer, Berlin, Heidelberg, Aug. 2002.

J. O. Ramsay. When the data are functions. *Psychometrika*, 47(4):379–396, 1982.

J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer series in statistics. Springer, New York, NY, second edition edition, 2005.

V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic learning in a random world*. Springer, New York, NY, 2005. ISBN 978-0-387-00152-4 978-0-387-25061-8. OCLC: 634599165.

H. Wang and S. J. Marron. Object oriented data analysis: Sets of trees. *The Annals of Statistics*, 35(5):1849–1873, 2007.

A. Willis. Confidence sets for phylogenetic trees. *Journal of the American Statistical Association*, 114(525):235–244, 2019.

G. Zeni, M. Fontana, and S. Vantini. Conformal Prediction: a Unified Review of Theory and New Challenges. *arXiv:2005.07972*, May 2020.

Appendix 1

Conformal prediction intervals for different number of observations: $|\mathcal{I}_1| + |\mathcal{I}_2| = 30, 130, 230$ and calibration always to show the discrete possible level of α. From the plots it is clear how the line tents to the theoretical line
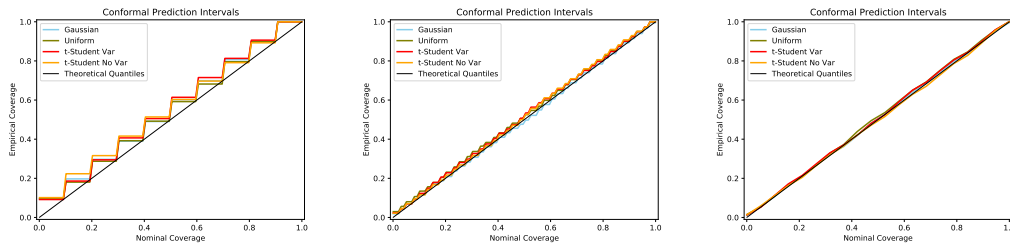
MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

41/2021 Costa, G., Cavinato, L., Maschi, C., Fiz, F., Sollini, M., Politi, L. S., Chiti, A., Balzarini, L., Ag
*Virtual Biopsy for Diagnosis of Chemotherapy-Associated Liver Injuries and
Steatohepatitis: A Combined Radiomic and Clinical Model in Patients with
Colorectal Liver Metastases*

40/2021 Martinolli, M.; Cornat, F.; Vergara, C.
*Computational Fluid-Structure Interaction Study of a new Wave Membrane
Blood Pump*

39/2021 Barnafi, N.; Di Gregorio, S.; Dede', L.; Zunino, P.; Vergara, C.; Quarteroni, A.
*A multiscale poromechanics model integratingmyocardial perfusion and
systemic circulation*

38/2021 Giusteri, G. G.; Miglio, E.; Parolini, N.; Penati, M.; Zambetti, R.
*Simulation of viscoelastic Cosserat rods based on the geometrically exact
dynamics of special Euclidean strands*

36/2021 Parolini, N.; Dede', L; Antonietti, P. F.; Ardenghi, G.; Manzoni, A.; Miglio, E.; Pugliese, A.; V
*SUIHTER: A new mathematical model for COVID-19. Application to the
analysis of the second epidemic outbreak in Italy*

37/2021 Dassi, F.; Fumagalli, A.; Mazzieri, I.; Scotti, A.; Vacca, G.
*A Virtual Element Method for the wave equation on curved edges in two
dimensions*

34/2021 Bonaventura, L.; Gatti F.; Menafoglio A.; Rossi D.; Brambilla D.; Papini M.; Longoni L.
An efficient and robust soil erosion model at the basin scale

35/2021 Regazzoni, F.; Quarteroni, A.
*Accelerating the convergence to a limit cycle in 3D cardiac
electromechanical simulations through a data-driven 0D emulator*

33/2021 Lupo Pasini, M.; Gabbi, V.; Yin, J.; Perotto, S.; Laanait, N.
*Scalable balanced training of conditional generative adversarial neural
networks on image data*

31/2021 Ferraccioli, F.; Arnone, E.; Finos, L.; Ramsay, J.O.; Sangalli, L.M.
Nonparametric density estimation over complicated domains