Conformal Prediction for Simulation Models

Benjamin LeRoy [1] **Chad Schafer** [1]

Abstract

Simulation models are often employed in situations where the relationship between a set of features x and an outcome y is too complicated to allow for the use of an analytical form. Examples are prevalent in the sciences, where the complex processes that generate observable data make explicitly learning the distribution of the data impossible, but theoretical understanding of the data generating process does allow for its simulation. In this paper we present a conformal approach to quantify uncertainty for situations where a simulator can generate potential values of y, the object of interest. Our approach combines split conformal inference to reduce computational costs and nested conformal inference to make sure our prediction regions can be explicitly defined even for complex outcome spaces. We provide strategies for setting tuning parameters, and demonstrate the approach on multiple examples including when y is a multivariate function.

1. Introduction

Often, the appropriate interpretation of predictions of future observations requires accurate quantification of uncertainty in those forecasts. Prediction regions, constructed to contain the true observation with some level of probability $1 - \alpha$, are commonly utilized. In order to achieve this coverage property (referred to as *validity*), different approaches resort to different types of assumptions (Bayesian priors, Gaussian and IID regression residuals, smoothness of the generating density, etc.) (Wasserman, 2006, pg. 6). All methods that create prediction regions have a vested interest in providing validity across as many settings as possible while also being *efficient*, i.e., limiting the size of the region.

This work focuses on creating prediction regions that achieve these dual goals through the use of a simulation model. In the canonical setup, we have a set of features X that are related to an outcome variable Y, and we have available a simulation model that generates potential y values given a particular value of $X = x$. We also assume that we have an available training sample of observed (x, y) pairs. Our objective is a prediction region for Y, hence we are more interested in the space that Y is from, which we denote $\mathcal{Y}$ and call the *outcome space*. The work in this paper focuses on building prediction regions for a wide range of outcome spaces, from the common Euclidean spaces to more complex geometric objects (e.g. the path of a hurricane or the impact of an epidemic on the proportion of a population infected versus susceptible over time). It is instructive to stress that our setup is not consistent with that used in *simulation-based inference* (e.g. Cranmer et al., 2020), as we are quantifying the uncertainty we have about y values with simulations of potential y values.

In the remainder of this paper we define $z_i = (x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$ as the fundamental unit of observation, and we assume that (for prediction purposes) we can observe x_i prior to y_i. We also assume that we have a simulation model that can produce sets of simulations in $\mathcal{Y}$ denoted $Sim(x_i, B) = \{\hat{y}_i^b\}_{b=1}^B \in \prod_{b=1}^B \mathcal{Y}$, with B being the number of simulations generated by the model.

To create prediction regions with valid coverage for a wide range of different outcome spaces our approach utilizes ideas from *conformal prediction* (also known as *conformal inference*). Conformal prediction provides a means to construct prediction regions with marginal finite sample validity while only assuming the exchangeability of observations (Vovk et al., 2005). With such a minimal assumption, conformal prediction has been applied to situations in which models can be viewed as "black boxes," including complex regression models, classification models, and in many other situations (e.g. respectively Lei et al., 2018; Shafer and Vovk, 2008; Lei and Wasserman, 2014). Nevertheless, the use of simulation models in conjunction with conformal prediction is novel.

Standard conformal prediction constructs a prediction region by assessing the *conformity* of every potential y to previously observed data. This is unrealistic in motivating cases where y is a geometric object, such as a tropical cyclone path, existing in an outcome space that cannot be completely explored. Through a combination of the *split*

[1]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. Correspondence to: Benjamin LeRoy <bpleroy@stat.cmu.edu>.

Submitted to the 2021 ICML workshop on distribution-free uncertainty quantification, Copyright 2021 by the authors.

conformal and *nested conformal* algorithms, we are able to to avoid training multiple simulation models and also the need to assess the conformity of every potential y value.

In the remainder of this paper we present our approach and discuss its properties and use. In Section 2, we introduce related work and the foundations of conformal inference, as well as discuss simulation-based prediction regions and prediction regions for complex functional data (a primary motivator for this work). In Section 3 we discuss our method (and variants) in detail with an exploration of a basic two-dimensional outcome space. In Section 4 we examine our proposed method's attributes applied to two *oracle*[1] examples; the first with a simple one-dimensional outcome space and the second with a more complex multivariate functional outcome space. In Section 5 we discuss overall observations and potential future work.

2. Related Work

This section highlights related and relevant work in conformal prediction, simulation/observation-based prediction regions, and prediction regions for complex functional data.

Conformal prediction defines finitely valid prediction regions while only assuming the exchangeability of observations, and hence is widely applicable (Vovk et al., 2005). Generally speaking, conformal prediction creates prediction regions by first defining an application-specific *conformal score* (cs) which measures to what extent a potential new observation y'_{new} conforms with our expectation of what it should be (based on its features x_{new} and previously observed data). Conformal prediction then defines the prediction region to contain all y'_{new} that have a conformal score greater than the $100 \cdot \alpha\%$ lowest conformal scores from previously observed data.

Split conformal prediction (Vovk et al., 2005, pg. 110-111) is a commonly-used procedure to apply conformal inference. This approach leverages sample splitting so that only a single model need be fitted, unlike the original *online* approach that allows for sequential analysis but requires the model to be refit for each of a new observation's potential values (see Vovk et al., 2005, pg. 17-27). Split conformal prediction is initiated by dividing the fully observed observations into a training and calibration set. The training set is used to construct a model, and then conformal scores are calculated for each observation in the calibration set relative to the trained model. For a new observation x_{new}, we build a prediction region from those y'_{new} that have conformal score greater than the α percentile of the calibration set's conformal scores. Algorithm 1 presents the decision protocol for including a potential y'_{new} into the prediction region under split conformal inference in the setting considered here.

Algorithm 1 Split Conformal Inference

Input: Nonconformal measure A, significance level ϵ, training examples $\{z_1, \ldots, z_m\}$, calibration examples $\{z_{m+1}, \ldots, z_n\}$, object x_{new} and a potential y'_{new}, where $z_i = (x_i, y_i)$
Task: to decide if y'_{new} should be in the prediction region $\Gamma^\epsilon(x_{new}; z_1, \ldots z_n)$.
Algorithm:

1. Provisionally set $z_{new} = (x_{new}, y'_{new})$

2.1. *Nonconformal scores for the calibration data*: For all $j \in \{m+1, \ldots, n\}$ set

$$\alpha_j = A(z_j; z_1, \ldots, z_m)$$

2.2. *Nonconformal scores for new observation*:

$$\alpha_{new} = A(z_{new}; z_1, \ldots, z_m)$$

3. Set

$$p_{y'_{new}} = \frac{\#\{j = m+1, \ldots, n, new \ : \alpha_j \geq \alpha_{new}\}}{n - m + 1}$$

4. Include y'_{new} in $\Gamma^\epsilon(x_{new}; z_1, \ldots z_n)$ if $p_{y'_{new}} > \epsilon$.

Conformal prediction regions can be viewed as level sets in the outcome space defined by the conformal score function. Specifically, the region contains all y'_{new} such that $cs(y'_{new}|x'_{new})$ is greater or equal to some threshold. Hence, a given conformal score function defines a collection of nested level sets (Vovk et al., 2005, pg. 9). Specifically, denote the level set as $\ell(\lambda) = \{y : cs(y|x) \geq \lambda\}$ then $\ell(\lambda) \subseteq \ell(\lambda')$ if $\lambda > \lambda'$.

Gupta et al. (2020)'s recent work on *nested conformal prediction* demonstrated the value of reversing the relationship between the conformal score and its associated nested level sets. Specifically, Gupta et al. (2020) showed that if one creates a sequence of nested sets, one could define a *radius function* that relates containment in the level sets to a *nonconformal*[2] *score*. This radius function is defined as

$$r(x, y) := \inf\{t \in \mathcal{T} : y \in \mathcal{F}_t(x)\} \ ,$$

where $\{\mathcal{F}_t(x)\}_{t \in \mathcal{T}}$ is a set of nested sets where $\mathcal{F}_t(x) \subseteq \mathcal{Y}$ and $\mathcal{F}_t(x) \subseteq \mathcal{F}_s(x)$ when $s \leq t$ and $\mathcal{T}$ captures the full sequence of t values for $\mathcal{F}_t$. Nested conformal prediction has the added benefit of naturally defining the shapes of the

[1] In this paper, we use *oracle* to imply that the simulations are drawn from the true conditional distribution.

[2] Conformal prediction also defines nonconformal scores which are the reverse of a conformal score - meaning the higher a the nonconformal score the more extreme the observation is.

prediction regions explicitly, which allows the user to avoid needing to evaluate all potential y'_{new} values. Prediction regions structures are often well defined (e.g. Lei et al., 2018), but this approach is very useful when they are not.

There is a long history of leveraging previous observations and/or simulations to create prediction and confidence regions. Possibly the most well known is bootstrap prediction and confidence regions, which can rely on either parametric assumptions or independence of residuals to define prediction bands that contain the desired level of probability (Wasserman, 2006, Chapter 8).

Transformation of simulated or observed data into a region of fixed probability is typically accomplished via quantile-based bands or density level sets. These approaches are optimally efficient in cases where the data distribution is known. In conformal inference, Romano et al. (2019) created quantile regression approaches that use conformal prediction and Lei and Wasserman (2014) has created conformal prediction regions from kernel density level sets based on a set of previously observed data. In this paper we follow behind Lei and Wasserman (2014) and aim to approximate density level sets relative to the simulations generated. Many papers have focused on optimal (and asymptotically consistent) ways to estimated a density's true level set while being computationally viable, include Walther (1997) and Baíllo et al. (2000). Both Walther (1997) and Baíllo et al. (2000) developed methods to estimate density level sets through different unions of balls around observed or simulated points, but with different geometric assumptions.

Our proposed method focuses on being used for simulations in both complex and simple outcome spaces. One example of a complex space where prediction regions are desirable but can be hard to create is when the outcomes are functional data (Ramsay and Silverman, 2005). Lei et al. (2015) and Fontana et al. (2020) have demonstrated how conformal prediction regions can be created for univariate functional data (with observations $f(\cdot)$ such that $f(t) \in \mathbb{R}^1$, and $t \in [0, 1]$). Fontana et al. (2020) constrained the prediction region by defining them relative to functional residuals. Lei et al. (2015) defined a prediction region through pointwise upper and lower points. Though Lei et al. (2015)'s approach was more flexible, it is only able to provide a conservative prediction region as they could not evaluate all potential outcomes with their conformal score - given the outcome space was infinite dimensional. Both of these functional papers also focus on univariate (one-dimensional) functional data and are not able to be extended to multivariate functional data (where the space the function projects into is multi-dimensional). We see our work as helping extend conformal prediction in complex application spaces such as these.

3. Methods

In this section we propose a split conformal approach for quantifying uncertainty that utilizes nested conformal prediction to define prediction regions for simple and complex outcome spaces $\mathcal{Y}$. Our method only requires that we possess a simulator that can generate a set $Sim(x_i, B)$, and that the outcome space be endowed with a distance metric, $d_{\mathcal{Y}}$. For illustration purposes, we will introduce the proposed method under the assumption that $\mathcal{Y} \subseteq \mathbb{R}^d$, and that the associated distance $d_{\mathcal{Y}}$ is Euclidean distance. This section will also discuss theoretical attributes of our method (Section 3.2) as well as desirable practical features (Section 3.3). We will show how practitioners can select associated tuning parameters in Section 3.4. Finally, in Section 3.5, we will propose extensions to our method to generalize to outcome spaces beyond $\mathbb{R}^d$.

3.1. Basic Method

Using nested conformal inference, as described in Section 2, we first define a procedure that takes in simulated points $(Sim(x_i, B))$ and returns a sequence of nested sets. The procedure starts by ranking all members of $(Sim(x_i, B))$. We propose ranking these simulations with a Gaussian kernel pseudo-density[3] estimator based on $d_{\mathcal{Y}}$, and $Sim(x_i, B)$ with values

$$\lambda_i^b = \frac{1}{B}\sum_{\ell=1}^{B} K\big(d_{\mathcal{Y}}(\hat{y}_i^b, \hat{y}_i^\ell)/\sigma\big) \tag{1}$$

for $b = 1, 2, \ldots, B$, and where

$$K(u) := \exp(-u^2)\,.$$

Note that for $\mathbb{R}^d$, these could be associate with a kernel density estimator if we scaled them by $1/((2\pi)^{d/2}\sigma)$. We can then define $\lambda_i^{(b)}$ to be the b^{th} order statistic of the kernel pseudo-density values. We construct our nested level sets $\{\mathcal{F}_t\}_{t=0}^{B}$ as follows:

1. $\mathcal{F}_0(Sim(x_i, B)) := \mathcal{Y}$ (the full space)
2. $\mathcal{F}_t(Sim(x_i, B)) := \bigcup_{\substack{\{\hat{y}_i^b \in Sim(x_i,B):\\ \lambda_i^b \geq \lambda_i^{(t)}\}}} \mathcal{B}(\hat{y}_i^b, \delta_i^{b,t})$

   for $t \in \{1, ..., B\}$, where $\mathcal{B}(\hat{y}_i^b, \delta_i^{b,t})$ is a ball centered at $\hat{y}_i^b$ with radius $\delta_i^{b,t}$.

This can seen as defining any particular set $\mathcal{F}_t$ as a union of balls around the subset of simulation points $\hat{y}_i^b \in$

[3]The pseudo-density defined below is not correctly scaled to be a density estimate, and pseudo-densities are more general as they can be defined on spaces that cannot have densities associated with them, see Ciollaro et al. (2016) for examples and more details.

$Sim(x_i, B)$ with $\lambda_i^b \geq \lambda_i^{(t)}$ with a *small-ball-radius*[4] $\delta_i^{b,t}$. In order for this sequence of sets to be nested ($\mathcal{F}_t \subseteq \mathcal{F}_{t-1}$), we require $\delta_i^{b,t} \leq \delta_i^{b,t+1}$.

We can then define a conformal score for these nested sets as the mapping

$$cs(y) = \max\{t \in \{0, ..., B\} : y \in \mathcal{F}_t\}\,. \tag{2}$$

Hence, if an observation y has a conformal score of t then $y \in \mathcal{F}_t$ and $y \notin \mathcal{F}_s$ for $s > t$. Figure 1 shows a sequence of nested sets for an example set of simulations where $\mathcal{Y} \subseteq \mathbb{R}^2$ and $\delta_\star^{b,t}$ is a fixed value (for all b and across values of t). Figure 1 displays three of the nested sets related to these simulations, and highlights the conformal score that a potential point y would have if it was contained in the observed level set, but not the level set created with one fewer simulation point. This figure also illustrates that the index t of the set $\mathcal{F}_t$ is directly related to the number of points used for the construction of the region.

3.2. Theoretical properties

This manner of defining a sequence of nested sets and associated conformal score (cs) has some desirable properties. First, our prediction regions are naturally conservatively valid. This follows from the well-known conformal conservative coverage guarantee, which is described for the split conformal version of nested conformal prediction in Gupta et al. (2020, pg. 6)'s Proposition 1. This can be rewritten in our notation as follows:

Proposition 1 *Suppose* $\{(X_i, Y_i)\}_{i \in [n] \cap \{n+1\}}$ *are exchangeable, and we have (i) disjoint training and calibration sets* $\mathcal{I}_1$ *and* $\mathcal{I}_2$*, (ii) conformal score* cs *as defined in equation 2, (iii) quantile function defined as*

$$Q_{1-\alpha}(cs, \mathcal{I}_2) := \lceil (\alpha)(|\mathcal{I}_2| + 1) \rceil / (|\mathcal{I}_2| + 1)\text{-quantile of } \{cs(y_i|x_i)\}_{i \in \mathcal{I}_2}\,,$$

and (iv) prediction set function $C(\cdot)$ *defined as*

$$C(X) := \mathcal{F}_{Q_{1-\alpha}(cs, \mathcal{I}_2)}\,.$$

Then

$$\mathbb{P}(Y_{n+1} \in C(X_{n+1}) | \{(X_i, Y_i)\}_{i \in \mathcal{I}_1}) \geq 1 - \alpha\,.$$

Below we define a modified nested set structure with an additional source of randomness τ, resulting in a new prediction set function $C'(X, \tau)$ and the approach being a smooth conformal predictor. This predictor will obtain the exact coverage guarantee as described in Vovk et al. (2005, pg. 27)'s Proposition 2.4.

[4]We use the phrase "small-ball-radius" to distinguish from the radius function of Gupta et al. (2020), but these are just gemetric radii, not necessarily small ones.

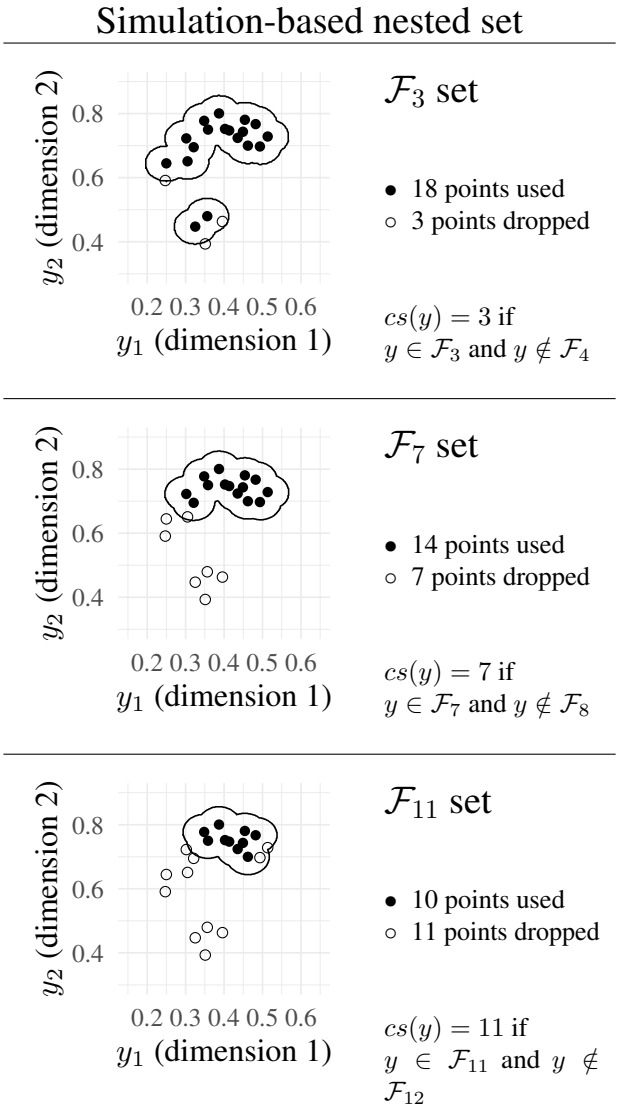

Figure 1. This figure presents a subset of the nested sets in $\{\mathcal{F}_t\}_{t=0}^{21}$ relative to a set of simulation points. As one moves from top to bottom the rows represent level sets with more an more simulation point removed for use to define the prediction region. We highlight the simulation points that are used to create the level set in solid black dots (that have high enough density estimates), and those that are not used in open circles. Notice how each of the regions is nested, and visually can be seen as have the same small-ball-radius across all level sets.

Proposition 2 *Define a function*

$$C'(X,\tau) := \begin{cases} \mathcal{F}_{b-1}(X) & \text{if } \tau < \tau^\star \\ \mathcal{F}_b(X) & \text{if } \tau \geq \tau^\star \end{cases}$$

where $b = Q_{1-\alpha}(cs, \mathcal{I}_2)$ *and*

$$\tau^\star = \frac{\alpha(|\mathcal{I}_2|+1) - |\{i = \mathcal{I}_2 : cs(y_i|x_i) > t\}|}{|\{i = \mathcal{I}_2 : cs(y_i|x_i) = t\}| + 1}. \quad (3)$$

Moreover, suppose for any new (X_{n+1}) *is paired with a* $\tau_{n+1} \sim Unif(0,1)$*, independent of* (X_{n+1}) *and* (Y_{n+1})*. Then the associated nested set conformal predictor is a smooth conformal predictor and following Vovk et al. (2005)'s Proposition 2.4, is exactly valid.*

Our proposed nested sets are similar to density level set estimates, which are the most efficient prediction regions under oracle settings (i.e., when the simulator draws from the true distribution) as discussed in Lei and Wasserman (2014). Although Section 3.4 will discuss multiple possibilities for defining $\delta_i^{b,t}$, our approaches mirror ideas in Baíllo et al. (2000), which selects the radius $\delta_i^{b,t}$ for a given level set by finding the minimum radius required to have every point covered by a union of balls around the other points. Baíllo et al. (2000) showed that this converges to the true level set for any fixed density cutoff.

Moreover, with sufficiently small $\delta_i^{b,t}$ and under the simulators' distribution conditional on x_i, the proportion of mass contained by the union of balls $\mathcal{F}_t$ would be close to t/B. In other words, if an observation has a conformal score of $\lfloor \gamma \cdot B \rfloor$, we would expect proportion γ of the simulations to be contained in the set that the observation was contained in. Under an oracle setting where the simulator captured the true generative distribution, $cs(Y|x)/B$ would be a discrete approximation of a highest (posterior) density region (HPD) value. In the oracle setting this can be seen as providing the most efficient prediction regions in conformal inference at a marginal and local level (Izbicki et al., 2021). We will examine this property in the examples in Section 4.

3.3. Practical properties

Beyond the theoretical properties, we want to highlight the definition of $\mathcal{F}_0$ and its potential relevance to the practitioner. If the simulation model underestimated the spread (uncertainty) of the true distribution one would expect a non-trivial proportion (ξ) of the calibration set to have a conformal scores of 0 (contained only in the $\mathcal{F}_0$, defined as the full response space). By construction, that would mean if the practitioner wished to have a prediction region with confidence greater than or equal to $(1-\xi)$, then this approach would predict $\mathcal{F}_0$, i.e. provide a non-informative response. Practitioners would be able to use this ξ to express the maximum level of coverage that could be given without being uninformative.

3.4. Selecting tuning parameters

This approach has three tuning parameters: the number of simulations, the σ value for the density estimate, and the $\delta_i^{b,t}$ that defines the small-ball-radii of the balls that comprise the sets. In this section we discuss the selection of the number of simulations, describe an approach to selecting σ, and a few ways to specify the $\delta_i^{b,t}$.

The number of simulations a model outputs is often a trade-off between the time and computational complexity to create and process the simulations and a desire to have enough simulations to well represent the simulated distribution. For our method, most of the computations can be performed in advance of obtaining new observations for which predictions are required. Ultimately, the general prediction region creation for a single observation is not too costly. Because of the discrete nature of the associated conformal score, more simulations will lead to a more granular conformal score distribution and a more granular prediction region. In Section 4 we vary the number of simulations, and show that smaller B does naturally lead to increased volatility. Ultimately, this choice will be subject to the computational resources available to the user.

The most important tuning parameter for this method is likely σ, the scaling for the distance based pseudo-density estimator. In Section 3.5 and Section 5 we will discuss other ways to order the simulations, but density-related ordering helps the regions inherit efficiency properties from density level sets. We propose tuning the σ value using methods similar to those used in mode clustering (e.g. Lei et al., 2013, Proposition 4.1). Given that we only assume access to a distance measure $d_{\mathcal{Y}}$, we propose using a grid search across the quantiles of the pairwise distances between simulations to select σ. We suggest using an associated quantile value (e.g. 25%) to then be used across different X values. To actually select the optimal quantile with grid search, we suggest the practitioner select a few (at least three) sets of simulations with different number of modes, and use the mean-shift mode clustering algorithm (e.g. Ciollaro et al., 2016) to cluster the observations with different values of σ-quantile values. The optimal σ value should then be selected if it identifies the correct number of mode clusters for each of the different simulation set examples.

A more complex parameter to tune is the small-ball-radii of the balls that make up the sequence of nested sets. We propose and compare a few approaches in Section 4. The first is to define a single small-ball-radius for every point and every set in the sequence of sets. We propose a rule of thumb of using the minimum covering radius on the top 80% of the observations relative to the density estimate. This is motivated by work by Baíllo et al. (2000) that use minimum covering radius to define a estimate of a density level set, and the 80% threshold mirrors suggestions in local depth

literature like (Agostinelli and Romanazzi, 2011, pg. 825)).

A more complex approach to define the small-ball-radii is to vary $\delta_i^{b,t}$ relative to the associated level set that is being estimated (i.e, for any point $\hat{y}_i^b$, $\delta_i^{b,t}$'s value varies as we vary t). We cannot strictly follow approaches such as Baíllo et al. (2000) or Walther (1997) as we also need to constrain the sets to be nested, but we propose a constrained approach that reflects Baíllo et al. (2000)'s structure. Specifically, we define a function

$$rad_{vary}(t) = \max_{b \leq t} \min_{\substack{b' \leq t \\ b' \neq b}} d(y_\star^{(b)}, y_\star^{(b')}) \ , \tag{4}$$

and then define the small-ball-radius $\delta_\star^{b,t}$ as

$$\delta_\star^{(b),t} = \max_{t \leq s \leq (b)} rad_{vary}(s) \ .$$

The second condition enforces the nested set constraint. For any fixed level set to estimate, and assuming a more discretized $rad_{vary}(t)$ function above that desired threshold, we can obtain convergence to the true level set as the number of simulations $B \to \infty$ which mirrors Baíllo et al. (2000).

Proposition 3 *For $\mathcal{Y} \subset \mathbb{R}^d$, compact. As the number of simulations n for any given x_i goes to infinite, let $rad_{vary}(t)$ vary only at a discrete number of t value associated with fixed density level set cutoffs $\{c_j\}_{j=1}^P$. Then each level set $\mathcal{F}_{t'}$, where t' is the number of simulations with density estimates greater or equal to c_j and with $\delta_{i,t}$ defined as above, will converge true density level set with cutoff c_j.*

Without this discretization of the $rad_{vary}(t)$, we observe volatility in $rad_{vary}(t)$, for which we suggest smoothing out the $rad_{vary}(t)$ function. This is examined in Section 4.2 as well.

Beyond just defining the small-ball-radii of the union of balls, we also explore if mode based structure can help meet containment criteria. This is specifically designed for more complex data (e.g. functional data) which might have different measures of containment (e.g. uniform containment versus pointwise). Defining conformal scores relative to containment in mode groups can naturally be combined with a the small-ball-radii selection approaches. Specifically, for each mode cluster of the simulations, treat the creation of the nested sets separately as if there was only 1 mode. This includes assessing containment of an observation; the observation must be completely contained in a single mode's structure to be defined as contained. To combine the nested sets of the multiple modes, we use the global pseudo-density estimates to order the simulations and defined the relative $\mathcal{F}_t$. More concretely, let $\mathcal{F}_t = \cup_{m \in \{1,\ldots,M\}} \mathcal{F}_{t_m}^m$ where $\mathcal{F}_{t_m}^m$ is the level set for mode m made with simulations from m mode cluster with simulations that have pseudo-densities greater or equal to the t^{th} lowest pseudo-density across all modes.

3.5. Extending to different outcome spaces

To extend to more complex spaces, including the example in Section 4.2, we envision that only one change is required to adapt to these spaces. Although the small-ball-radii can be estimated relative to any the distance function, in new spaces we can image the prediction region construction may have to change a little for the best interpretive value. For example, Section 4.2 defines our nested prediction regions for multivariate functions as regions relative to the function's own multidimensional output space. We imagine this is very space dependent, but that only minor changes in assessing containment and the geometric definitions of the prediction regions.

4. Examples

To assess our new method, we demonstrate two different uses under oracle settings (i.e., with simulators that draw from the true conditional distribution). The first has an outcome space in $\mathbb{R}^1$ which we use to compare against other conformal approaches, and examine the effect of the number of simulations B. Our second example demonstrates the versatility of our proposed approach by using a more geometrically complex outcome space.

4.1. Example with $\mathbb{R}^1$ Outcome Space

In our first example, we utilize a generative distribution from Lei and Wasserman (2014), defined by the following set of equations:

$$\begin{aligned} X &\sim \text{Unif}(-1.5, 1.5) \\ (Y|X=x) &\sim 0.5N(f(x) - g(x), \sigma^2(x)) \\ &\quad + 0.5N(f(x) + g(x), \sigma^2(x)) \ , \end{aligned} \tag{5}$$

where

$$\begin{aligned} f(x) &= (x-1)^2(x+1) \\ g(x) &= 2\sqrt{(x+0.5)} \cdot \mathbb{I}(x \geq -0.05) \\ \sigma^2(x) &= 1/4 + |x| \ , \end{aligned}$$

and is visualized in Figure 2.

To assess the properties of our method, we construct a calibration set consisting of 300 observations. Our simulation model, being an oracle, returns draws from the true conditional distribution. We compare the effect of using $B = 200$ and $B = 1000$ simulations to define the nested sets for both calibration and test observations. Given the simplicity of this example's outcome space, we will only demonstrate

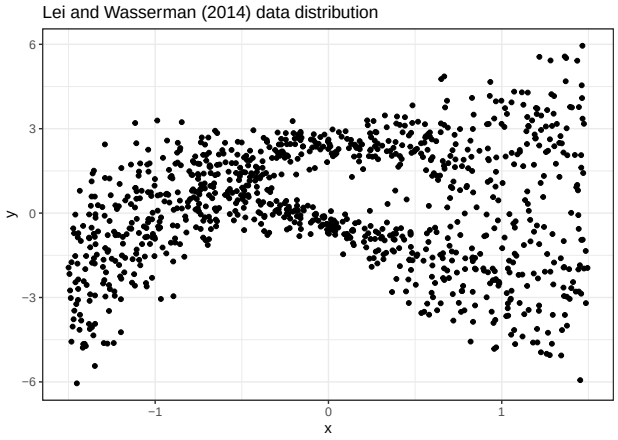

Figure 2. Visualization of the distribution of data generated from Equation 5 first proposed by Lei and Wasserman (2014).

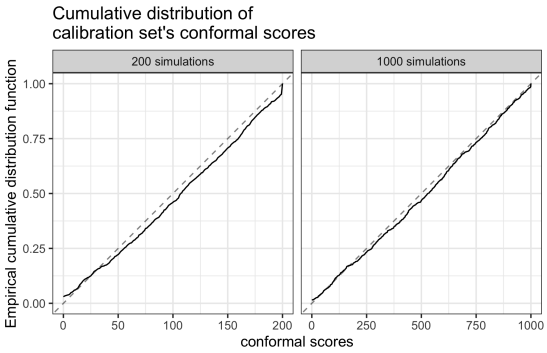

Figure 3. Visualizations of cumulative distribution conformal scores for the calibration set for the first example for $B = 200$ and $B = 1000$ simulations. A dashed grey line is included to show the optimal cumulative distribution if the conformal scores were distributed normally. One can see the empirical cumulative distribution varies more from the dashed line with $B = 200$ than $B = 1000$, but that overall both empirical cumulative distributions suggest that the conformal scores are roughly uniformly distributed.

our approach with a fixed radius (defined by the minimum covering distance for the top 80% of the simulations) and will use σ-quantile of 25%.

Figure 3 visualizes the distribution of conformal scores from the calibration set for both $B = 200$ and $B = 1000$. As discussed at the end of Section 3, in an oracle setting we would hope that each nested set $\mathcal{F}_t$ would contain approximately t/B observations. Hence a natural diagnostic of our proposed method would be to see if the distributions of the conformal scores are roughly uniform in this case, which seems to be the case for the conformal scores for both $B = 200$ and 1000 in Figure 3.

An additional means of assessing our method's performance is to examine the empirical local coverage of the prediction regions. Given the oracle setting and the similarity of our method's conformal scores to high posterior density (HPD) values, we expect our approach to outperform other local conformal approaches that use CDE values as conformal scores. We define a test set with X values that vary across a fine grid between -1.5 and 1.5. We set the desired confidence level $(1 - \alpha)$ to be 0.6 for the clearest visual comparison. For each x value, we create prediction regions based on $B = 200$ and $B = 1000$ simulations. To visualize each region we took a fine grid of potential y values between -8.5 and 8.5, and determined which of these were contained in the prediction region. For each x value we also generated 300 new samples from the true distribution of Y conditioned on $X = x$ and examined the proportion of these samples that were contained in the prediction region. The results of these two assessments can be see in the subplots of Figure 4a and 4b. The prediction regions fit the data very well, and across each value of X we observe that approximately 60% of the newly created test observations are covered. We can observe that when moving from $B = 200$ simulations (Figure 4a) to $B = 1000$ simulations (Figure 4b), we see a decrease in the variability of the regions' shape (relative to what is expected) and a decrease in the variability of the containment proportions.

We compare our method to other approaches that utilize the conditional density estimates and HPD values as conformal scores. For each of the following approaches, we use split conformal inference with 300 points in the calibration set, and with a conditional density estimate fixed equal to the truth. The simplest approach (the *global CDE approach*) uses the conditional density estimate as the conformal score. The second approach (the *global HPD approach*) uses HPD values as the conformal score, and has been proposed more recently in works such as Gupta et al. (2020, Appendix D) and Izbicki et al. (2021). Third, we again use the conditional density estimate as the conformal score, but follow Lei and Wasserman (2014)'s *local conformal* approach. Local conformal approaches partition $\mathcal{X}$ and obtain provide finite-sample validity per partition. Following the approach of Lei and Wasserman (2014), we partition the $\mathcal{X}$ space into eight bins and for an individual test x value we use calibration conformal scores (CDE values) only from those calibration points with x values in the same bin as the test x value. Our fourth approach again uses the conditional density estimate as the conformal score but follows Izbicki et al. (2021)'s `CD-split+` local approach. This approach is similar to Lei and Wasserman (2014)'s approach, except in that it groups the x values relative to their predicted conditional density estimates using a *profile distance*. (See Izbicki et al.

(2021) for more details on the profile distance.) We assess these procedures in the same fashion as described to assess ours.

Visualization of this assessment for the above procedures can be found in the c, d, e, and f subplots of Figure 4 (respectively relative to presentation order in the above paragraph). Under these oracle conditions the global HPD and our approach preform almost identically. Compared to the CDE based approaches (even those designed for localized validity), our method outperforms them in obtaining exact local validity. Note that all approaches are conservatively marginally valid by construction.

4.2. Example with geometrically complex outcome space

One of the promising features of our proposed conformal approach, as applied to simulation models, is that the extension to more complicated spaces is straightforward. In this subsection we present such an example, modeling the evolution of epidemics as multivariate functional data. We define a hierarchical generative process as follows:

$$\begin{aligned} X &\sim \text{Unif}(0,1) \\ (X'|X=x) &= 2\cdot(x-0.5) \\ (Z|X'=x') &\sim 0.5\cdot N(f(x')-g(x'),\sigma^2(x')) \\ &\quad + 0.5\cdot N(f(x')+g(2\cdot(x'),\sigma^2(x')) \\ (R_0|Z=z) &= 1/7\cdot z+2 \\ (f|R_0=r_0) &\sim SIR(\beta=0.1,\gamma=0.1/r_0) \end{aligned} \tag{6}$$

where

$$\begin{aligned} f(x') &= (x'-1)^2(x'+1) \\ g(x') &= 2\sqrt{x'+0.5}\cdot\mathbb{I}(x'\geq -0.05) \\ \sigma^2(x') &= 1/4+|x'|\ . \end{aligned}$$

The object of interest is the complete trajectory of the epidemic defined by the path of the epidemic through a three-dimensional unit simplex, where a point represents the proportion of the population that is susceptible, infected, or recovered at a given time. In the generative process, we use an approximation to the *susceptible, infected or recovered (SIR) model* in the form of a Bernoulli discrete approximation of the standard ODE SIR model, with 1000 individuals of which 50 are initially infected (Kermack and McKendrick, 1927; Tuckwell and Williams, 2007). Recent work in epidemiology such as Gallagher (2019) encourages the assessment of this trajectory in a time-invariant manner, and we do so by treating each epidemic trajectory as a geometric filament. For this example we used the R package, `EpiCompare`, to simulate SIR epidemics (Gallagher and LeRoy, 2021).

The hierarchical generative process is visualized in Figure 5. The top subplot highlights the distribution of the latent variable R_0 relative to the value of x, and includes a red line to highlight the distribution of R_0 values when $x = 0.64$. The bottom subplot presents a visual representation of simulated epidemic filaments in the three-dimensional unit simplex (representing the proportion of the population susceptible, infected or recovered at a given time) associated with an x value of 0.64.

Prior to the creation of prediction regions we need to make two decisions. First, we define a filament distance between two epidemics as

$$d_{\mathcal{Y}}(f_1,f_2) := \int_{t=0}^{1}(f_1(t)-f_2(t))^2dt\ . \tag{7}$$

In this distance we view each filament f as a function defined on $t\in[0,1]$, where, for example, $f(0.25)$ is the point that is a quarter of the length of the filament away from its starting point. Second, since we are interested in predicting where the true epidemic trajectory will fall, we define each set $\mathcal{F}_t$ as the union of small filamental bands around each simulation with a psuedo-density estimate greater or equal to the t^{th} lowest psuedo-density estimate. We also define the minimum covering distance of an simulated epidemic (used to determine $\delta_i^{b,t}$) as the smallest width of the filamental bands centered other simulations to fully cover the simulated epidemic.

4.2.1. Selecting tuning parameters

In order to demonstrate the selection of σ, we first select three X values that have different modal and spread structure. For this example we create simulations for $X\in\{0,0.64,0.85\}$ as the first has simulations with one mode and the later have simulations with two modes. Additionally, $X=0.64$ and $X=0.85$ have different "spreads" of their simulations. We select fixed randomization seeds so that the simulations from these X values have the expected number of modes in a visual comparison and then run mode clustering on the simulations across a grid of potential σ-quantiles (see Ciollaro et al., 2016, for examples.). For $B=200$ we find σ-quantiles of 15%-25% identified the correct mode structure and for $B=1000$, we found the σ-quantiles 25%-30% performed optimally. We selected to use σ-quantiles at 20% and 25% for $B=200$ and 1000 respectively. This selection also included selecting mode clustering parameters associated with convergence of the algorithm (see Fukunaga and Hostetler (1975) for more details).

Given the complexity of the output space and the volatility of our varying radius approach, we will also use $X\in\{0,0.64,0.85\}$ to identity potential smoothings of the rad_{vary} function (from equation 4). To do so we first

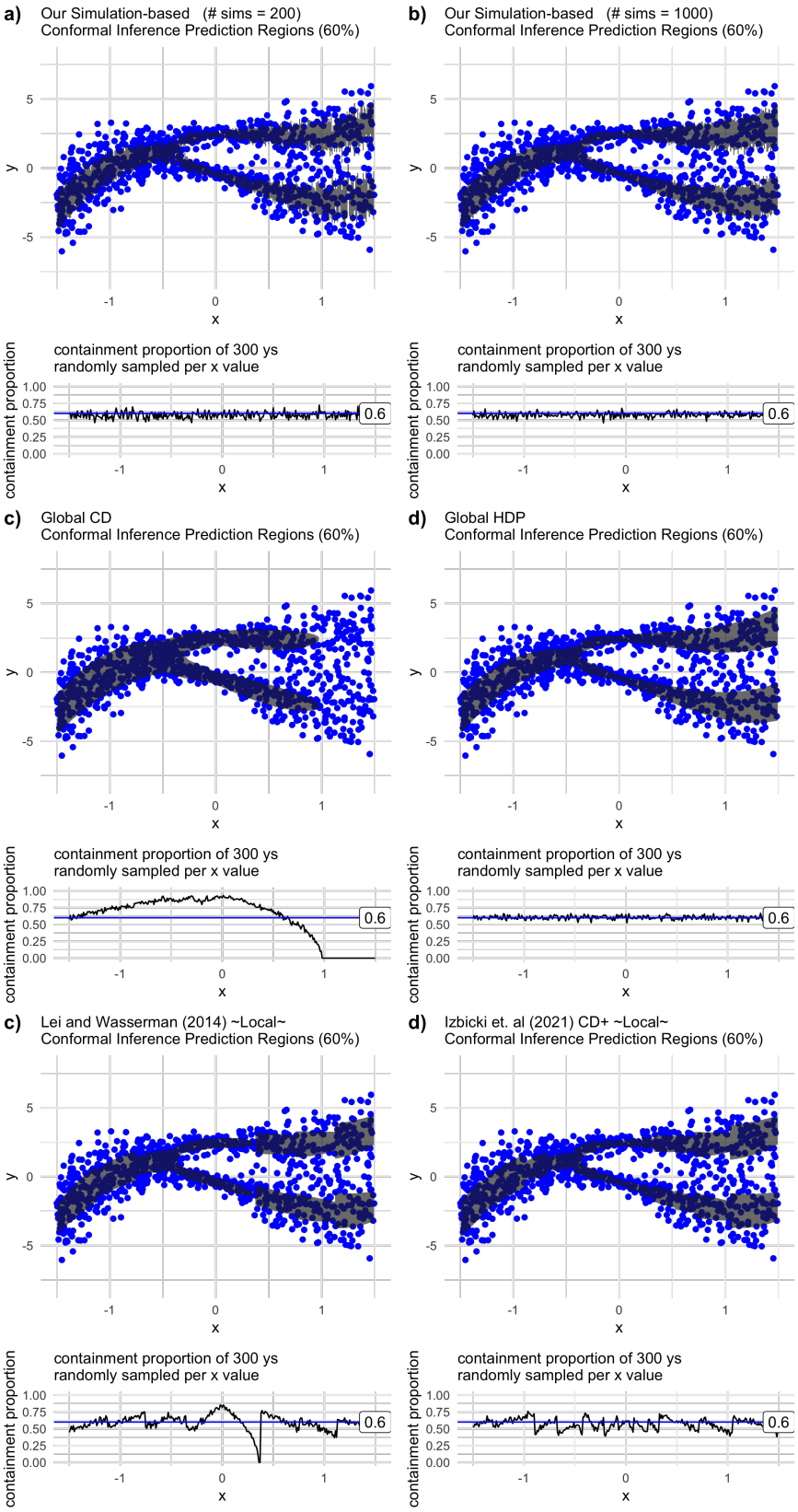

Figure 4. A comparison of different conformal inference approaches and how each approach perform on a local / conditional level under an oracle fit. From **a)** to **f)**, we include our simulation-based approach with $B = 200$, our simulation-based approach with $B = 1000$, standard conformal approach using the CDE as the conformal score, standard conformal approach using HDP as the conformal score, Lei and Wasserman (2014)'s local conformal approach using the CDE as a conformal score, and Izbicki et al. (2021)'s `CD-split+` local conformal approach using the CDE as a conformal score. For each approach we examine prediction regions across the range of potential x values both visually (in the larger subplot) and estimated the coverage of the prediction region using 300 new samples generated from the true distribution of Y conditioned on $X = x$.

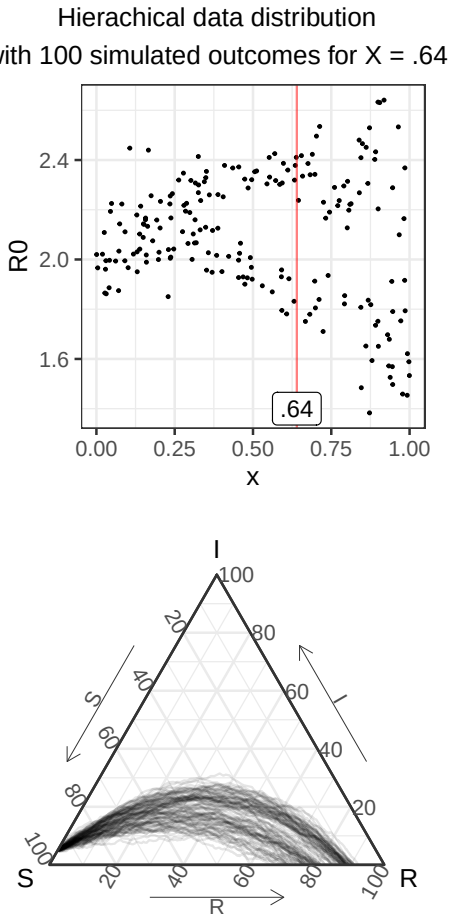

Figure 5. Visualizing the generative distribution as described in Equation 6, where the right plot is a sampling of epidemics where the associated x value is 0.64.

define eight smoothing functions, four of which are smoothing splines with varying degrees of freedom $(50, 25, 10, 5)$, and the other four as a moving minimum function with varying window sizes $(5, 10, 25, 50)$. To select the optimal smoothing functions we create prediction regions based on B simulations for the three X values, and then create B more simulations and examine how closely the distribution of conformal scores of the second set are to a uniform distribution. We compare to a uniform distribution because we designed the conformal scores to mirror the structure of HPD values, which are uniformly distributed. We then collect KL distances between the empirical and expected distribution for 20 runs (per X and B values) and visually select the optimal smooth function. For both $B = 200$ and 1000, we selected a smoothing spline with with 5 degrees of freedom, and for $B = 200$ we also select a moving minimum with length 25 whereas for $B = 1000$ we select a moving minimum function with window length of 50.

4.2.2. Application of Approach

Based on the selected tuning parameters, we will look at prediction regions constructed with (i) fixed radius, (ii) fixed radius with mode structure, (iii) varying radius, and (iv) varying radius with mode structure. We also examine prediction region approaches with smoothed varying radius approaches, smoothed either with a smoothing spline or with a moving minimum function (with parameters specified above). Similar to Section 4.1, we use 300 points in a calibration set.

First we consider the distribution of conformal scores based on different nested set constructions. Figure 6 visualizes these distributions for $B = 1000$ across all different predictor set constructors. In each facet, we visualize a histogram of the conformal scores as well as a empirical cumulative distribution. If the conformal scores were uniformly disturbed we would observe the cumulative distribution close to the grey line (which has a slope of $300/1000$). The facet columns are based on whether the set creation used mode structure as described at the end of Section 3.5, and the rows capture different approaches to specifying the radius. In this example, the fixed radius approaches get closest to a uniform distribution (with the approach using mode clustering getting slightly closer). The conformal scores that result from the unsmoothed varying radius approach skew towards higher values. Although both smoothing approaches reduce this skew, the distributions of the conformal scores for all varying radius approaches are highly concentrated on a few values around 1000 (for $B = 1000$). This is less than ideal as this means $70\%, 80\%, 90\%$ confidence regions may all appear identical since the quantile of the set of conformal scores is the same.

We now pivot to examining local performance. As with the example in Section 4.1, we applied our approach across a gird of test X values between 0 and 1. We examined the containment of the prediction regions on 300 simulations draw from the conditional distribution. Figure 7 presents the proportion of the simulations contained in prediction regions with $40\%, 60\%$ and 80% confidence. The fixed radius approaches appear to consistently provide desirable local coverage with only minor differences between the bimodal and unimodal R_0 groups ($X < .25$, $X \geq .25$). All varying radius approaches are very conservative at the 80% confidence level, capturing 100% of the simulations. Smooth approaches seem to not be as conservative at the lower confidence levels but are more volatile in the empirical coverage amounts than the fixed radius approaches and seem more effected by the change of the number of modes. Coverage performance does not seem to vary much depending on whether or not we use mode structure.

To understand why our varying radius methods are under performing in this oracle setting, we select a single X value

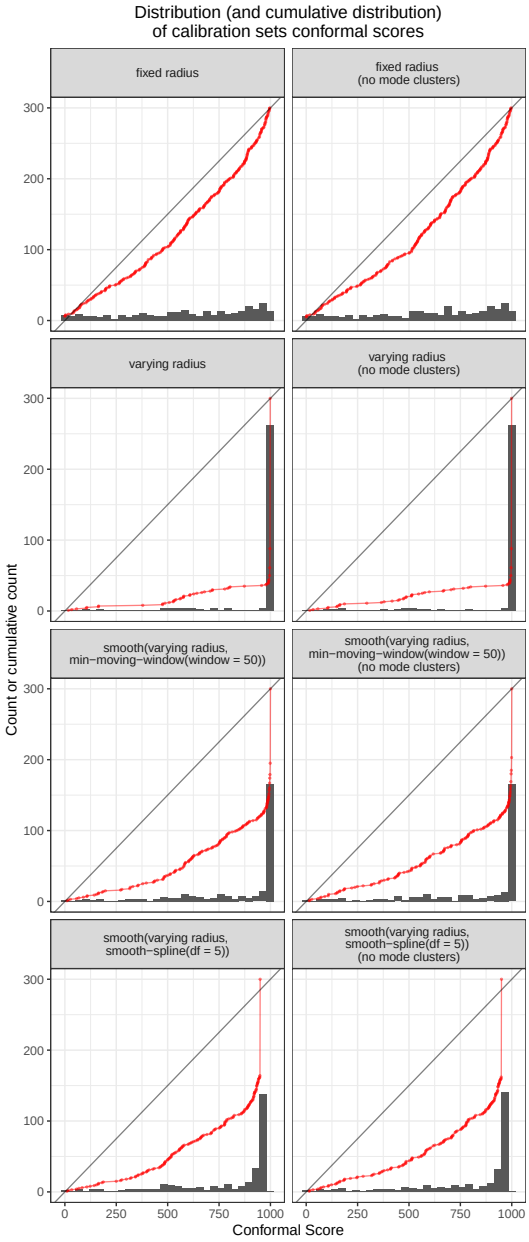

Figure 6. For $B = 1000$ simulations, this figure displays the 300 calibration samples' conformal scores across different nested level/radii approaches. We examine the histogram and cumulative distribution of these scores for each approach.

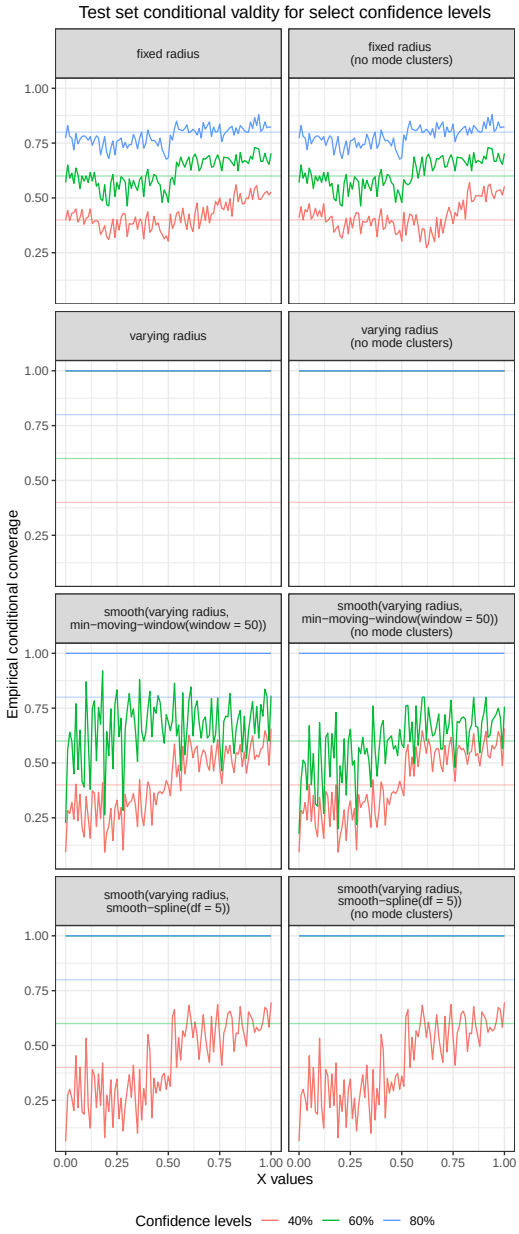

Figure 7. For $B = 1000$ simulations, this figure presents the empirical proportion of new test simulations contained given then are associated with a specific X value and we examine across a grid of X values. We examine this proportion for confidence levels 40%, 60% and 80% and across the multiple radii defining approaches. The lighter colored line is the expected proportion and the darker line is the empirical coverage.

($X = 0.8$) and visualize the sequences of nested sets for each construction approach in Figure 9. Figure 8 provide information about the simulation set structure for $X = 0.8$ as well. In Figure 8 notice that the lower cluster of simulations tend to have lower pseudo densities. For the fixed radius, we can see that there are a large number of distinct prediction regions captured by the small squares varying in color in the fixed radius subplots in Figure 9. On the other hand, for the non-smoothed varying radius prediction regions it appears that there are only a few different level sets. It is possible that this was caused by a very large value from the rad_{vary} function early on that lead to an early $\mathcal{F}_t$ (with

high t) individual simulations all having a large radius. This would led to the appearance that there was only two discrete level sets, which matches well with the patterns observed in Figure 7.

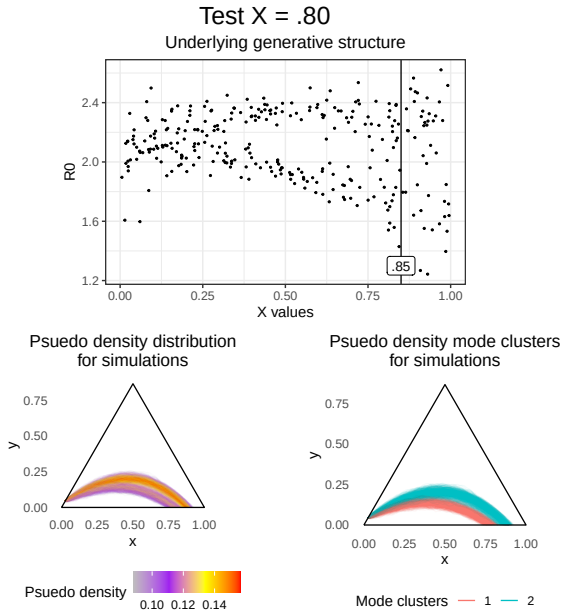

Figure 8. For $B = 1000$, this figure captures the underlying structure that all prediction regions have access to and the where along the generative distribution where $X = .8$ belongs. The lower two subplots highlight the psuedo-density values for the simulated curves and which mode cluster they below to.

Generally speaking, this example shows the potential usefulness of our approach on complex outcome spaces, but also the potential danger in utilizing varying radius approaches as currently proposed.

5. Discussion and Future Work

In this discussion we will make comparisons between our work and current work in the literature, discuss potential alterations to our approach, and present a few avenues of future work.

The objective of this work is to bring conformal inference to simulation-based models. We also show how the use of flexibility of simulation models in conjunction with nested conformal inference can bring conformal inference to outcome spaces of greater complexity. Our two examples show the potential of our approach to do just that, but also highlight different complexities that a practitioner would need to address.

It is important to consider comparisons between our approach and existing conformal tools. For outcomes spaces

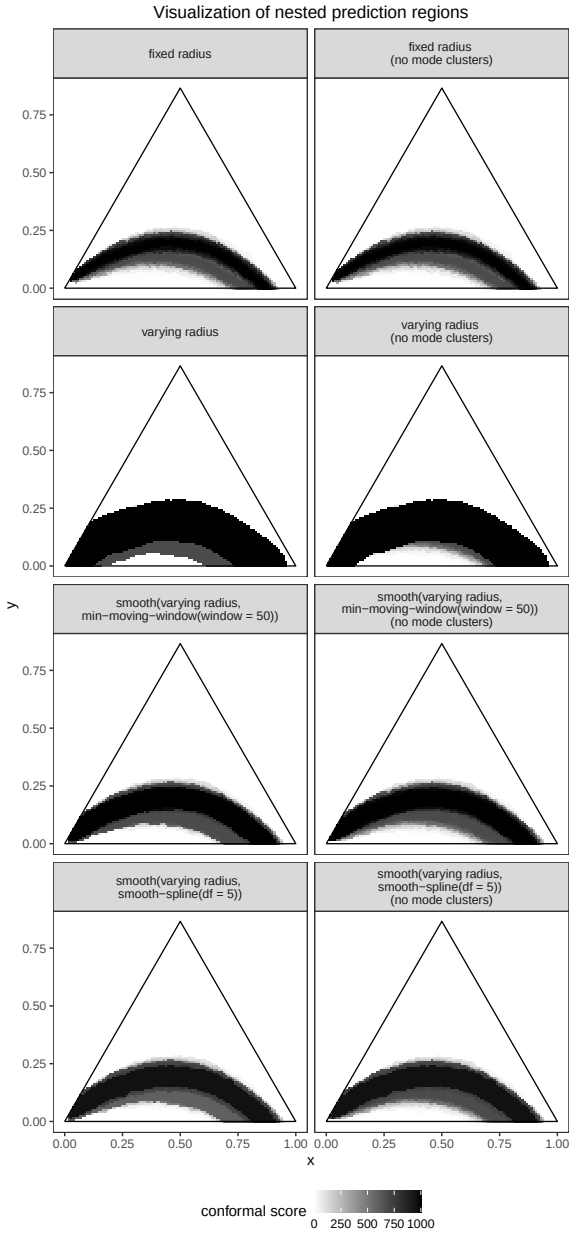

Figure 9. For $B = 1000$ simulations, this figure visualizes a the sequence of nested sets for each radii approach for $X = .8$. Figure 8 provides supplementary information about what these approaches use to create the prediction regions. Notice that there only seems to be a small sequence of nested sets for the varying radius approaches. This is likely due to large early values of rad_{vary} that made an early nested level set very large. Problems like this can't occur with the fixed radius approaches, and results in being able to see many different nested level sets.

in $\mathbb{R}^d$, Lei and Wasserman (2014) and Izbicki et al. (2021)'s `HPD-split` use conditional density estimates or HPD values as conformal scores. From a theoretical view with well

fit conditional density estimates, using HPD values as conformal scores lead to desired local coverage and density level sets produce efficient prediction regions. Moreover, in one-dimensional outcomes spaces, these approaches would naturally outperform simulation models (if the base CDE model and the simulation models reflected the same distribution), due to the smoothness of the implicit nested set structure created by a CDE or HPD function. However, when the outcome space is $\mathbb{R}^d$, $d > 1$, HPD and estimating density level sets become more computationally expensive, in part because computational implementations of HPD often require having a high resolution CDE grid for the outcome space. As such, we suggest that our approach may provide a useful avenue for high-dimensional Euclidean outcome spaces, as most of our structure relies on distances between simulations which are less impacted by the dimension of the space than high resolution grids.

Moving beyond Euclidean outcome spaces, in Section 2 we discussed two conformal approaches that extended conformal inference to one-dimensional functional data. Lei et al. (2015) helped motivate this work, and our approach appears visually very similar in applications with a one-dimensional functional data outcome space (specifically, $f(t) \in \mathbb{R}^1, t \in [0, 1]$). In particular given the use of mode clustering in Lei et al. (2015), practitioners might be hard-pressed to justify using our approach. Nevertheless, our approach provides an avenue to extending to functional data with $f(t) \in \mathbb{R}^d$ and $t \in \mathbb{R}^d$. Extensions of both Lei et al. (2015) and Romano et al. (2019) (see Section 2 for refreshers) to higher dimensional functional data are not intrinsically natural and the most natural extension might be with convex containment regions (which are less efficient than ones demonstrated in Section 4.2). The work presented here represents a useful contribution to quantifying uncertainty in such situations.

Our proposed approach requires multiple decisions, each of which affects the resulting prediction region. These include the selection of the ranking function (λ_i^b). We settled on using a pseudo-density estimate given the associated efficiency of prediction regions in Euclidean outcome spaces built using density estimates. Other approaches to ranking, including depth and local depth (e.g. Geenens and Nieto-Reyes, 2017; Paindaveine and Van bever, 2013), could be an interesting approach for practitioners in non-Euclidean outcome spaces. Use of depth to calculate ranks would lead to prediction regions that appear more like the intersection of the depth level set and α-hull (Krasnoshchekov and Polishchuk, 2014) around the top simulations. This is less interpretable in terms of notions of depth, but by construction still obtains marginal finite-sample validity if included in our pipeline.

The other decision a practitioner faces is the means of constructing the level sets. We used unions of balls with different radius selection tools, and both examples suggest that varying radius approaches preformed worse than the more simple fixed radius approaches. Nonetheless we see these approaches as well motivated by arguments in Baíllo et al. (2000) and Walther (1997). A geometric analysis that attempts to define a sequence of nested sets that converge to density level sets may naturally define a better way to create nested sets than a fixed radius approach. We can also imagine that model clustering (which did not seem to affect the strength of the method in our examples) might be useful in other settings and can be made more computationally efficient through the use of distance-centric clustering or by combining projection approaches (e.g. Lee et al., 2008) and mode clustering in the projection space.

Finally, we see two avenues for future work. The first is application to real world examples, with imperfect simulation models. Ongoing work will apply this approach to tropical cyclone feature modeling, with the end goal of informing human-in-the-loop processes. A second objective is to show how our approach can be combined with other ideas in conformal inference, in particular (Izbicki et al., 2021)'s `CD-split+` local conformal approach, which clustered the $\mathcal{X}$ space using estimated HPD values from a conditional density estimate. We envision that, given our conformal score similarity to HPD, that `CD-split+` might be extended to work with our approach.

6. Acknowledgements

This work is supported by the NSF AI Institute: Physics of the Future, NSF PHY-2020295. Some of the simulations for in this work were carried out on the Vera system at the Pittsburgh Supercomputing Center in collaboration with the McWilliams Center for Cosmology.

References

Claudio Agostinelli and Mario Romanazzi. Local depth. *Journal of Statistical Planning and Inference*, 141(2): 817–830, feb 2011. ISSN 03783758. doi: 10.1016/j.jspi.2010.08.001.

Amparo Baíllo, Antonio Cuevas, and Ana Justel. Set Estimation and Nonparametric Detection. *The Candian Journal of Statistics / La Revue Canadienne de Statistique*, 28(4): 765–782, 2000. ISSN 1098-6596.

Mattia Ciollaro, Christopher R. Genovese, and Daren Wang. Nonparametric clustering of functional data using pseudo-densities. *Electronic Journal of Statistics*, 10(2):2922–2972, 2016. ISSN 19357524. doi: 10.1214/16-EJS1198.

Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of*

the National Academy of Sciences of the United States of America, 117(48):30055–30062, 2020. ISSN 10916490. doi: 10.1073/pnas.1912789117.

Matteo Fontana, Simone Vantini, Massimo Tavoni, and Alexander Gammerman. A Conformal Approach for Distribution-free Prediction of Functional Data. In Germán Aneiros, Ivana Horová, Marie Hušková, and Philippe Vieu, editors, *Functional and High-Dimensional Statistics and Related Fields*, pages 83–90. Springer International Publishing, 2020. ISBN 9783030477561. doi: 10.1007/978-3-030-47756-1.

Keinosuke Fukunaga and Larry D. Hostetler. The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition. *IEEE Transactions on Information Theory*, 21(1):32–40, 1975. ISSN 15579654. doi: 10.1109/TIT.1975.1055330.

Shannon Gallagher and Benjamin LeRoy. Time invariant analysis of epidemics with EpiCompare. 2021. URL https://github.com/skgallagher/EpiCompare/.

Shannon K Gallagher. *Catalyst: Agents of change; Integration of compartment and agent-based models for use in infectious disease epidemiology*. PhD thesis, Carnegie Mellon University, 2019.

Gery Geenens and Alicia Nieto-Reyes. On the functional distance-based depth, 2017.

Chirag Gupta, Arun K. Kuchibhotla, and Aaditya K. Ramdas. Nested Conformal Prediction and quantile out-of-bag ensemble models. may 2020. URL http://arxiv.org/abs/1910.10562.

Rafael Izbicki, Gilson Shimizu, and Rafael B. Stern. CD-split and HPD-split: efficient conformal regions in high dimensions. *arXiv*, pages 1–31, 2021. ISSN 23318422.

William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

Dmitry Krasnoshchekov and Valentin Polishchuk. Order-k α-hulls and α-shapes. *Information Processing Letters*, 114(1-2):76–83, 2014. ISSN 00200190. doi: 10.1016/j.ipl.2013.07.023.

Ann B. Lee, Boaz Nadler, and Larry Wasserman. Treelets—An adaptive multi-scale basis for sparse unordered data. *The Annals of Applied Statistics*, 2(2), jun 2008. ISSN 1932-6157. doi: 10.1214/07-AOAS137.

Jing Lei and Larry Wasserman. Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 76 (1):71–96, 2014. ISSN 13697412. doi: 10.1111/rssb.12021.

Jing Lei, James Robins, and Larry Wasserman. Distribution-free prediction sets. *Journal of the American Statistical Association*, 108(501):278–287, 2013. ISSN 01621459. doi: 10.1080/01621459.2012.751873.

Jing Lei, Alessandro Rinaldo, and Larry Wasserman. A conformal prediction approach to explore functional data. *Annals of Mathematics and Artificial Intelligence*, 74 (1-2):29–43, 2015. ISSN 10122443. doi: 10.1007/s10472-013-9366-6.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. Distribution-Free Predictive Inference for Regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018. ISSN 1537274X. doi: 10.1080/01621459.2017.1307116.

Davy Paindaveine and Germain Van bever. From Depth to Local Depth: A Focus on Centrality. *Journal of the American Statistical Association*, 108(503):1105–1119, sep 2013. ISSN 0162-1459. doi: 10.1080/01621459.2013.813390. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.2013.813390.

James Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer New York, New York, NY, 2005. ISBN 9781461271666.

Yaniv Romano, Evan Patterson, and Emmanuel J Candès. Conformalized Quantile Regression. Technical report, 2019. URL https://github.com/yromano/cqr.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9: 371–421, jun 2008. ISSN 15324435. URL http://arxiv.org/abs/0706.3188.

Henry C. Tuckwell and Ruth J. Williams. Some properties of a simple stochastic epidemic model of SIR type. *Mathematical Biosciences*, 208(1):76–97, 2007. ISSN 00255564. doi: 10.1016/j.mbs.2006.09.018.

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer Science & Business Media, 2005. ISBN 0-387-00152-2.

Guenther Walther. Granulometric smoothing. *Annals of Statistics*, 25(6):2273–2299, 1997. ISSN 00905364. doi: 10.1214/aos/1030741072.

Larry Wasserman. *All of Nonparametric Statistics*. 2006. ISBN 9780387251455. doi: 10.1007/0-387-30623-4.