Exact and Approximate Conformal Inference for Multi-Output Regression

Chancellor Johnstone CHANCELLOR.JOHNSTONE@GEAEROSPACE.COM
Air Force Institute of Technology[*]
Eugene Ndiaye E_NDIAYE@APPLE.COM
Georgia Institute of Technology[†]

Editor: Khuong An Nguyen, Zhiyuan Luo, Harris Papadopoulos, Tuwe Löfström, Lars Carlsson and Henrik Boström

Abstract

It is common in machine learning to estimate a response y given covariate information x. However, these predictions alone do not quantify any uncertainty associated with said predictions. One way to overcome this deficiency is with conformal inference methods, which construct a set containing the unobserved response with a prescribed probability. Unfortunately, even with a one-dimensional response, conformal inference is computationally expensive despite recent encouraging advances. In this paper, we explore multi-output regression, delivering exact derivations of conformal inference p-values when the predictive model can be described as a linear function of y. Additional contributions include 1) a multivariate extension of an existing root-based method for conformal prediction (`rootCP`) and 2) the introduction of union-based approximations for conformal prediction (`unionCP`). Both of these methods provide efficient ways of approximating the conformal prediction region for a wide array of multi-output predictors, both linear and nonlinear, while preserving computational advantages. We also provide both theoretical and empirical evidence of the effectiveness of our methods using both real-world and simulated data.

Keywords: Uncertainty quantification, Nonlinear models, Prediction, Homotopy, Multi-task, Multivariate regression.

1. Introduction

In regression, we aim to predict (or estimate) some response y given covariate information x. These predictions alone deliver no information related to the uncertainty associated with the unobserved response, and thus, would benefit from the inclusion of a set $\Gamma^{(\alpha)}(x)$ such that, for any significance level $\alpha \in (0, 1)$,

$$\mathbb{P}\big(y \in \Gamma^{(\alpha)}(x)\big) \approx 1 - \alpha. \tag{1}$$

One method to generate $\Gamma^{(\alpha)}(x)$ is through conformal inference, which generates *conservative* prediction sets for some unobserved response y under only the assumption of exchangeability. Given a finite number of observations $\mathcal{D}_n = \{(x_i, y_i)\}_{i=1}^n$ and a new unlabelled example x_{n+1}, conformal prediction regions are generated through the repeated inversion of the test,

$$H_0 : y_{n+1} = z \quad \text{vs.} \quad H_a : y_{n+1} \neq z, \tag{2}$$

[*] Currently at GE Aerospace.
[†] Currently at Apple.

where y_{n+1} is the unobserved response for observation $n+1$ and z is a candidate response value (Lei et al., 2018). With conformal inference, a p-value for the hypothesis test shown in Equation 2 can be constructed by using an augmented dataset $\mathcal{D}_n \cup (x_{n+1}, z)$ and comparing one's ability to predict the new response y_{n+1} to the already observed responses. A *conformal prediction set* is the collection of candidates z for which the null hypothesis is not rejected, *i.e.,* when the error in predicting z is not too high compared to others.

The inversion of the test in Equation 2 is traditionally called "full" conformal prediction since it uses the entire dataset to learn a predictive model. Unfortunately, full conformal prediction is computationally demanding in many cases, with each new candidate point z requiring a new model to be fit. To avoid this complexity, more efficient methods, *e.g.,* split conformal inference (Vovk et al., 2005; Lei et al., 2018) and trimmed conformal inference (Chen et al., 2016), have been introduced, with trade-offs between computational efficiency and performance.

Of interest to our work in this paper are *exact* and *approximate* conformal inference methods, which aim to reduce computational complexity without sacrificing performance. Nouretdinov et al. (2001) showed that with certain models, ridge regressors in particular, conformity scores for every observation in a dataset can be constructed as an affine function of the candidate value z and only require training the model once.

We extend the result of Nouretdinov et al. (2001) to predictors of the form

$$\hat{y} = Hy, \tag{3}$$

where y is an $n \times 1$ vector of responses, H is an $n \times n$ matrix, and $\hat{y}$ is an $n \times 1$ vector of predictions. We note that H can also be a function of a set of covariates, *e.g.,* as with ridge regression, where $H = X(X^\top X + \lambda I)^{-1} X^\top$. In reality, the restriction shown in Equation 3 is more general than ridge regression; we only require the predictions be linear functions of the input. In this paper, we refer to models that follow Equation 3 as *linear* models; this is in contrast to the traditional usage of the term to reflect models that are linear with respect to their parameters.

There also exist methods for the efficient construction of a conformal prediction set through the use of root-finding procedures. As one example, Ndiaye and Takeuchi (2021) introduces the `rootCP` algorithm, which utilizes a traditional root-finding approach, *i.e.,* a bisection search, to find points on the boundary of a conformal prediction set with fewer model trainings than full conformal prediction.

In more complex settings it might be of interest to construct a model for multiple responses, *i.e.,* for some response y with support $\mathcal{Y} \subset \mathbb{R}^q$, also known as multi-output (or multi-task) regression (Zhang and Yang, 2018; Borchani et al., 2015; Xu et al., 2019). Thus, we might wish to construct a prediction set such that some q-dimension version of y, say $y = (y^{(1)}, \ldots, y^{(q)})^\top$, is contained with some specified probability.

Contributions With these potential scenarios is mind, we aim to extend exact conformal inference to the multi-output regression setting and subsume our contribution as

- an extension of exact conformal p-values to multiple dimensions with various predictors and conformity measures

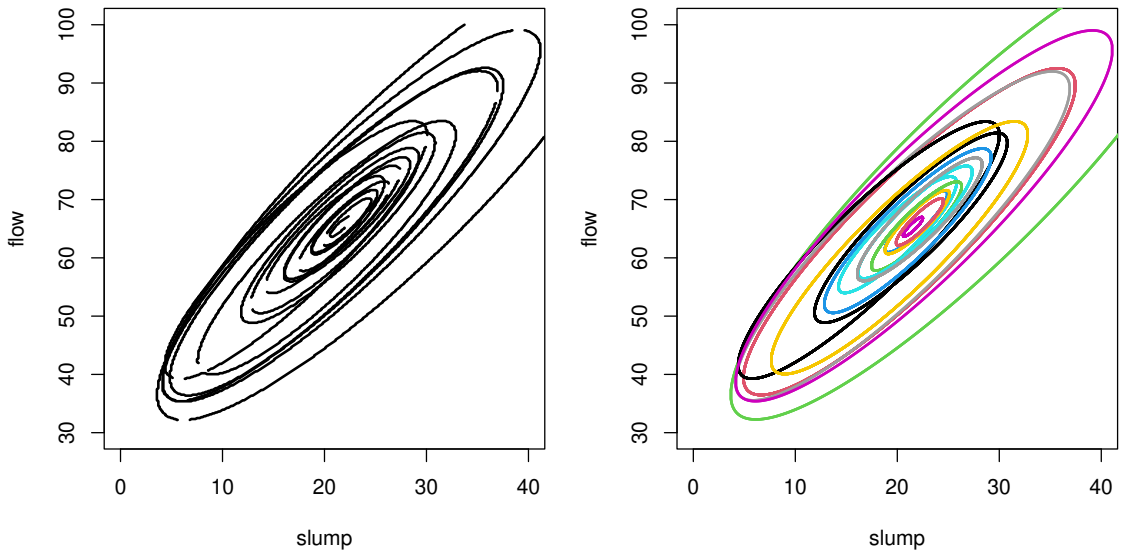


Figure 1: Comparing `gridCP` contours (left) to p-value change-point sets (right) constructed using $||\cdot||_W^2$ with $W = \hat{\Sigma}^{-1}$ for an observation from `cement` dataset. Each of the various colors in the right plot correspond to the different p-value change-point sets for each observation.

- an exact approach for sampling points on the boundary of a conformal prediction set and a multivariate extension of `rootCP`, originally introduced in Ndiaye and Takeuchi (2021).
- an approximate approach for conservative conformal prediction sets through `unionCP`.

The extension of exact conformal inference to multiple dimensions makes use of *p-value change-point sets* (described in detail in Section 3) which allow for the description of conformal prediction regions. The extension of `rootCP` reduces the trade-off between various conformal inference methods, balancing the computational efficiency of split conformal prediction (`splitCP`) with the performance of full conformal prediction (`fullCP`).

Table 1 summarizes the overall computational costs for each of these methods in terms of the number of model retraining iterations required to generate the conformal prediction region. We also include the computation complexity of the CP approximation provided by `gridCP`, where a finite grid of candidate points are tested to determine a prediction set. In contrast, `fullCP` comprises approaches where the exact conformal prediction set can be constructed in a closed-form. We include an example of p-value change-point sets compared to `gridCP` sets in Figure 1.

In the linear case, each of the methods for prediction set generation require the same number of model refits as `splitCP`. From Table 1, we can see that this is not the case of nonlinear models. We note that this does not account for the complexity of interval construction in each case. The rest of the paper is laid out as follows. Section 2 provides requisite background for the paper. Section 3 extends exact conformal inference to multiple dimensions delivers exact conformal inference results for several classical predictors. Section

Table 1: Computational complexity of methods where q is the response dimension, n is the number of observations, r is the cardinality of the candidate value set, d is the number of search directions, and ϵ is the tolerance.

Method	Nonlinear
`splitCP`	$\mathcal{O}(q)$
`gridCP`	$\mathcal{O}(rq)$
`fullCP`	-
`rootCP`	$\mathcal{O}\big(dq \log_2(1/\epsilon)\big)$

4 discusses several approaches to generate exact and approximate conformal prediction sets. Section 5 provides an empirical evaluation. Section 6 concludes the paper.

Notation We denote the $n \times 1$ design matrix as $X = (x_1, \dots, x_n)^\top$ and the $n \times q$ matrix of responses as $y = (y^{(1)}, \dots, y^{(q)})^\top$, where $y^{(k)}$ is the vector of responses for the k-th dimension. We define $\mathcal{D}_n = \{(x_i, y_i)\}_{i=1}^n$ as a collection of n observations, where the i-th data tuple (x_i, y_i) is made up of a covariate vector x_i and a q-dimensioned response y_i, where $y_i = (y_i^{(1)}, \dots, y_i^{(q)})$. The augmented dataset is denoted by $\mathcal{D}_{n+1}(z) = \mathcal{D}_n \cup \{(x_{n+1}, z)\}$. We denote the response matrix augmented with z as $y(z)$, where the candidate value *vector* is defined as $z = (z_1, \dots, z_q)^\top$. Given $j \in [n]$, the rank of an element u_j among a sequence $\{u_1, \dots, u_n\}$ is defined as $\mathrm{Rank}(u_j) = \sum_{i=1}^n \mathbb{1}_{u_i \leq u_j}$.

The main approach to perform the test inversion associated with Equation 2 relies on the Quantile Lemma (Tibshirani et al., 2019), which we include in this work (in a slightly different form) as Lemma 1

Lemma 1 *Let $U_1, \dots, U_n, U_{n+1}$ be an exchangeable sequence of random variables. Then, for any $\alpha \in (0, 1)$,*

$$\mathbb{P}\big(\mathrm{Rank}\,(U_{n+1}) \leq \lceil (1-\alpha)(n+1) \rceil\big) \geq 1 - \alpha.$$

One might consider a counter-example to Lemma 1 where $U_1 = \ldots = U_{n+1}$ w.p. 1. We can solve this issue by adjusting the rank construction to include a tie-breaking component τ_i where $\tau_i \sim U(0, 1)$ (Vovk et al., 2005; Lei et al., 2018).

2. Conformal Inference

Originally introduced in Gammerman et al. (1998) as "transductive inference", conformal inference (CI) was originally focused on providing inference with classification approaches. Vovk et al. (2005) provides a formalized introduction to conformal inference within regression. With the express purpose of inference, the goal of CI is to attach, in some fashion, a measure of uncertainty to a predictor, specifically through the construction of a conservative prediction set, *i.e.,* one such that

$$\mathbb{P}\big(y_{n+1} \in \Gamma^{(\alpha)}(x_{n+1})\big) \geq 1 - \alpha. \tag{4}$$

We wish to construct a prediction set for a new observation (x_{n+1}, y_{n+1}), where x_{n+1} is some known covariate vector and y_{n+1} is some, yet-to-be-observed response. Assuming each data pair (x_i, y_i) and (x_{n+1}, y_{n+1}) are drawn exchangeably from some distribution $\mathcal{P}$,

conformal inference generates conservative, finite sample-valid prediction sets in a distribution-free manner. In a prediction setting, test inversion for a particular candidate value z is achieved by training the model of interest on an augmented data set $\mathcal{D}_{n+1}(z)$. At this point, we leave our model of interest general, denoting the prediction of the i-th observation based on a model trained with $\mathcal{D}_{n+1}(z)$ as $\hat{y}_i(z)$. Following the refitting, each observation in the augmented data set receives a (non)conformity *measure*, which determines the level of (non)conformity between itself and other observations. One popular, and particularly effective, conformity measure is the absolute residual

$$S_i(z) = |y_i - \hat{y}_i(z)|. \tag{5}$$

We can construct the conformity *score* associated with a candidate point z as

$$\pi(z) = \frac{\text{Rank}(S_{n+1}(z))}{n+1}, \tag{6}$$

where $S_i(z)$ is the conformity measure for the data pair (x_i, y_i) as a function of z and $S_{n+1}(z)$ is the conformity measure associated with (x_{n+1}, z). Then, a p-value for the test shown in Equation 2 can be constructed as $1 - \pi(z)$. In Section 3 we explicitly describe the p-value associated with some z in terms of *p-value change-point sets*, which explicitly define where changes in rank occur between the conformity scores for specific observations. A prediction set for an unknown response y_{n+1} associated with some covariate vector x_{n+1} is

$$\Gamma^{(\alpha)}(x_{n+1}) = \{z : (n+1)\pi(z) \leq \lceil (1-\alpha)(n+1) \rceil\}. \tag{7}$$

Then, subuniformity holds for $(n+1)\pi(y_{n+1}) = \text{Rank}(S_{n+1}(y_{n+1}))$, and Equation 4 holds for $\Gamma^{(\alpha)}(x_{n+1})$. By the previous results, CI can also be utilized in the multivariate response case, where one is interested in quantifying uncertainty with respect to the joint behavior of a collection of responses, given a set of covariates. Thus, we can construct a multidimensional prediction set $\Gamma^{(\alpha)}(x_{n+1}) \subset \mathbb{R}^q$ such that Equation 4 holds when y_{n+1} is some q-dimensional random vector.

The initial extension of conformal inference to the multivariate setting comes from Lei et al. (2015), which applies conformal inference to functional data, providing bounds associated with prediction "bands". Diquigiovanni et al. (2022) extends and generalizes additional results for conformal inference on functional data. Joint conformal prediction sets outside the functional data setting are explored in Kuleshov et al. (2018) and Neeven and Smirnov (2018). Messoudi et al. (2020, 2021) provide extensions to these works through the use of Bonferroni- and copula-based conformal inference, respectively. Cella and Martin (2020), Kuchibhotla (2020) and Johnstone and Cox (2021) construct joint conformal sets through the use of depth measures, *e.g.,* half-space and Mahalanobis depth, as the overall conformity measure. Messoudi et al. (2022) extends these works by generating adaptive conformal predictive regions in multiple dimensions. Applications of conformal inference have been seen in healthcare (Olsson et al., 2022), drug discovery (Cortés-Ciriano and Bender, 2019; Eklund et al., 2015; Alvarsson et al., 2021), and decision support (Wasilefsky et al., 2023), to name a few. For a thorough treatment on conformal inference in general, we point the interested reader to Fontana et al. (2023) and Angelopoulos et al. (2023). We also point the reader to Hallin et al. (2010) for an approach to generate quantiles for multi-output regression.

2.1. Computationally Efficient Conformal Inference

Due to the inherent model refitting required to generate prediction sets through full conformal inference, *i.e.,* the testing of an infinite amount candidate points, more computationally efficient methods have been explored. We describe a subset of these methods in the following sections. Specifically, we focus on resampling-based and exact conformal inference.

Resampling Methods Split conformal inference (Vovk et al., 2005; Lei et al., 2018) generates conservative prediction intervals under the same assumptions of exchangeability as *full* conformal inference. However, instead of refitting a model for each new candidate value, split conformal inference utilizes a randomly selected partition of $\mathcal{D}_n$, which includes a training set $\mathcal{I}_1$ and a calibration set $\mathcal{I}_2$. First, a prediction model is fit using $\mathcal{I}_1$. Then, conformity measures are generated using out-of-sample predictions for observations in $\mathcal{I}_2$. The split conformal prediction interval for an incoming (x_{n+1}, y_{n+1}), when using the absolute residual as our comformity measure, is

$$\Gamma^{(\alpha)}_{\texttt{split}}(x_{n+1}) = [\hat{y}_{n+1} - s, \hat{y}_{n+1} + s], \tag{8}$$

where $\hat{y}_{n+1}$ is the prediction for y_{n+1} generated using the observations in $\mathcal{I}_1$, and s is the $\lceil(|\mathcal{I}_2| + 1)(1 - \alpha)\rceil$-th largest conformity measure for observations in $\mathcal{I}_2$. In order to combat the larger widths and high variance associated with split conformal intervals, cross-validation (CV) approaches to conformal inference have also been implemented. The first CV approach was introduced in Vovk (2015) as cross-conformal inference with the goal to "smooth" inductive conformity scores across multiple folds. Aggregated conformal predictors Carlsson et al. (2014) generalize cross-conformal predictors, constructing prediction intervals through any exhangeable resampling method, *e.g.,* bootstrap resampling. Other resampling-based conformal predictors also include CV+ and jackknife+ (Barber et al., 2021). For a more detailed review and empirical comparison of resampling-based conformal inference methods, we point the interested reader to Contarino et al. (2022).

Exact Conformal Inference for Piecewise Linear Estimators In order to test a particular set of candidate values for inclusion in $\Gamma^{(\alpha)}(x_{n+1})$, we must compare the conformity measure associated with our candidate data point to the conformity measures of our training data. Naively, this requires the refitting of our model for each new candidate value. However, Nouretdinov et al. (2001) showed that $S_i(z)$, constructed using Equation 5 in conjunction with a ridge regressor, varies piecewise-linearly as a function of the candidate value z, eliminating the need to test a dense set of candidate points through model refitting. Other exact conformal inference methods include conformal inference through homotopy (Lei, 2019; Ndiaye and Takeuchi, 2019), influence functions (Bhatt et al., 2021; Cherubin et al., 2021), and root-finding approaches (Ndiaye and Takeuchi, 2021). While not exact, Ndiaye (2022) provide approximations to the full conformal prediction region through stability-based approaches.

3. Conformal Inference for Multi-Output Regression

In the following sections, we extend the results in Nouretdinov et al. (2001) to multiple dimensions. While CI can be applied to any prediction or classification task, in this section

we restrict each of our predictors, given an incoming observation (x_{n+1}, z), to the form

$$\hat{y}_i^{(k)}(z_k) = \big(H_k(x_{n+1}, x_i)y^{(k)}(z_k)\big)_i, \tag{9}$$

where $\hat{y}_i^{(k)}(z_k)$ is the prediction for the i-th observation of the k-th response dimension, z_k is the k-th element of z, and $(\cdot)_i$ is the i-th element of the vector argument. H_k is an $(n+1) \times (n+1)$ matrix where each element is a function of $x_1, \ldots, x_{n+1}$. While H does explicitly depend on $x_1, \ldots, x_{n+1}$, we forgo including this dependency for notational convenience. We note that the restriction shown in Equation 9 is analogous to the restriction identified in Equation 3. We also note that while we define H_k in Equation 9 as a function of both x_i and the augmented point x_{n+1}, this definition is general enough so as to include many classes of predictors. As an example, we can describe H_k with respect to the k-th response dimension for ridge regression as

$$H_k(x_{n+1}, x_i) \equiv H_k(x_{n+1}) = \tilde{X}(\tilde{X}^\top \tilde{X} + \lambda_k I)^{-1}\tilde{X}^\top. \tag{10}$$

where $\tilde{X} = (x_1, \ldots, x_n, x_{n+1})^\top$ is the augmented design matrix. Additionally, we require that H_k be constructed independently of $y^{(k)}$, *i.e.,* not as a function of $y^{(k)}$. One focus of our paper is construction of exact p-values for a given z without retraining our model. We also identify how we construct explicit *p-value change-point sets*, denoted as $\mathcal{E}_i$ for the i-th observation, where

$$\mathcal{E}_i \equiv \{z \in \mathbb{R}^q : S_{n+1}(z) \leq S_i(z)\}. \tag{11}$$

We note that $\mathcal{E}_{n+1} \equiv \mathbb{R}^q$. Then, the p-value associated with the hypothesis test shown in Equation 2 for any candidate point z is

$$p\text{-value}(z) = \frac{|\{i \in [n+1] : z \in \mathcal{E}_i\}|}{n+1}. \tag{12}$$

The result of Nouretdinov et al. (2001) was extended to include both lasso and elastic net regressors in Lei (2019). For this paper, we utilize a generalized version, shown in Proposition 2.

Proposition 2 *Assume the fitted model as in Equation 3, where $H \equiv H(x_{n+1}, x_i)$. Then, if we define $y(z) = (y^\top, z)^\top$, we can describe the vector of residuals associated with the augmented dataset and some candidate value z as*

$$y(z) - Hy(z) = A - Bz$$

where $A = \big(I - H\big)y(0)$ and $B = \big(I - H\big)(0, \ldots, 0, 1)^\top$.

With Proposition 2, we can then describe the conformity measure for the i-th observation, when using Equation 5, as $S_i(z) = |a_i + b_i z|$, where a_i and b_i and the i-th elements of A and B, respectively.

In the following sections, we describe exact conformal inference results for conformity measure constructions using $||\cdot||_W^2$, as well as results for finding points on the boundary of a conformal prediction set for any conformity measure. We include discussion on exact results using ℓ_1 as a conformity measure in the extended version of this paper.

3.1. Exact p-values with $||\cdot||_W^2$

In this section, we consider conformity measures of the form

$$S_i(z) = r_i(z)^\top W r_i(z) = ||r_i(z)||_W^2, \tag{13}$$

where $r_i(z) = y_i - \hat{y}_i(z)$, and W is some $q \times q$ matrix. For this construction of conformity score, Proposition 3 shows that S_i becomes quadratic with respect to z.

Proposition 3 *Assume the fitted model $\hat{y}^{(k)}(z_k) = H_k(x_{n+1}, x_i)y^{(k)}(z_k)$ for each response dimension $k \in [q]$. Then, using Equation 13,*

$$S_i(z) = \begin{bmatrix} a_{1i} + b_{1i}z_1 \\ \vdots \\ a_{qi} + b_{qi}z_q \end{bmatrix}^\top W \begin{bmatrix} a_{1i} + b_{1i}z_1 \\ \vdots \\ a_{qi} + b_{qi}z_q \end{bmatrix}$$

where $a_i = (a_{1i}, \ldots, a_{qi})^\top$, $b_i = (b_{1i}, \ldots, b_{qi})^\top$, and a_{ki} and b_{ki} are the i-th elements of the vectors A_k and B_k, respectively, defined as

$$A_k = \big(I - H_k(x_{n+1}, x_i)\big)y^{(k)}(0) \text{ and } B_k = \big(I - H_k(x_{n+1}, x_i)\big)(0, \ldots, 0, 1)^\top. \tag{14}$$

In order to maintain the probabalistic guarantees inherent to conformal inference, we require W to be constructed exchangeably. Two constructions that satisfy exchangeability are: 1) W constructed independently of $\mathcal{D}_{n+1}(z)$, or 2) W constructed using all observations within $\mathcal{D}_{n+1}(z)$. However, in practice, setting $W = \hat{\Sigma}^{-1}$, the observed inverse-covariance matrix associated with the residuals from our q responses using a model constructed using only $\mathcal{D}_n$, performs well. The p-value associated with some z using sets constructed using Equation 13 is the same as in Equation 12. While Proposition 3 does not restrict the structure of W, limiting W to be a symmetric, positive semi-definite matrix ensures that the set $\mathcal{E}_i$ is not only convex, but ellipsoidal. Without this additional restriction on the matrix W, the p-value change-point sets could be ill-formed, *i.e.,* non-convex. For clarity, we describe how each $\mathcal{E}_i$ can be constructed in practice when using $||\cdot||_W^2$ by describing these regions in terms of conic sections in Section 3.1.1.

3.1.1. Change-point Sets as Conic Sections

With Proposition 3, $S_{n+1}(z) - S_i(z)$ is the difference between two quadratic forms. Thus, we can describe the boundary of $\mathcal{E}_i$ for every $i \in [n]$ as a *conic section*. Specifically, we can describe the difference between the candidate conformity measure and the conformity measure for observation i as,

$$S_{n+1}(z) - S_i(z) = V_{n+1}^\top W V_{n+1} - V_i^\top W V_i. \tag{15}$$

where $V_i = (a_{1i} + b_{1i}z_1, \ldots, a_{qi} + b_{qi}z_q)^\top$.

Knowing we aim to find the boundary of each $\mathcal{E}_i$, *i.e.,* the roots of Equation 15, we can expand the statement into the form of a conic section such that

$$(1, z_1, \ldots, z_q)[M_{n+1} - M_i](1, z_1, \ldots, z_q)^\top = 0, \tag{16}$$

where M_j, for any $j \in [n+1]$, is

$$M_j = \begin{bmatrix} a_{1j} & b_{1j} & & 0 \\ \vdots & & \ddots & \\ a_{qj} & 0 & & b_{qj} \end{bmatrix} W \begin{bmatrix} a_{1j} & b_{1j} & & 0 \\ \vdots & & \ddots & \\ a_{qj} & 0 & & b_{qj} \end{bmatrix}^\top \tag{17}$$

with a_{jk} and b_{jk} previously. If we additionally define the following:

- $(\cdot)_{qq}$: the lower $q \times q$ submatrix of the matrix argument
- $(\cdot)_{i,-j}$: the i-th row of the matrix argument, without the j-th element of that row
- L_i: the upper-triangular Cholesky matrix of $(M_i^*)_{qq}$ where $M_i^* = M_{n+1} - M_i$

with,

$$z_i^c = (M_i^*)_{qq}^{-1}(-M_i^*)_{1,-1} \text{ and } K_i = \frac{-det(M_i^*)}{det\big((M_i^*)_{qq}\big)}, \tag{18}$$

we can translate any point s on the unit-ball to a point z on the boundary of $\mathcal{E}_i$ with

$$z = \sqrt{K_i} L_i s + z_i^c.$$

3.2. Application to Some Classical Predictors

Many regression methods generate predictions that follow Equation 9. In the following section describe some results for local-constant and local linear regression.

Local Constant (Nadaraya-Watson) Regression Kernel regression (Nadaraya, 1964; Watson, 1964) is a nonparametric regression technique that utilizes kernel density estimators (Parzen, 1962). Using some kernel, *e.g.,* a Gaussian kernel, we can perform "local-constant" regression by using $H_k(x_{n+1}, x_i) \equiv H_k(x_{n+1}) = (w_1, \ldots, w_{n+1})^\top$ where each w_i is a vector of the normalized kernel values $K_h(\cdot)$ for each observation x_j centered on x_i, *i.e.,*

$$w_i = \big(K_h(x_i - x_1), \ldots, K_h(x_i - x_{n+1})\big)^\top \big(I_{n+1} \sum_{j=1}^{n+1} K_h(x_i - x_j)\big)^{-1}.$$

where K_h is our multivariate kernel of choice. In this work, we utilize a *product* kernel (Scott, 1992).

Local Linear Regression In contrast to the local-constant regression with the Nadaraya-Watson estimator, local-linear regression (Fan, 1992) utilizes a weighted version of the design matrix. Local-linear regression constructs an estimate for y_i, as a function of the candidate value pair (x_{n+1}, z), by using an adjusted design matrix,

$$\tilde{X}_i = \begin{bmatrix} 1 & (x_1 - x_i)^\top \\ \vdots & \vdots \\ 1 & (x_{n+1} - x_i)^\top \end{bmatrix},$$

and a diagonalized version of w_i, *i.e.,* $G(x_i) = diag(w_i)$, resulting in an H_k matrix such that

$$H_k(x_{n+1}, x_i) = \tilde{X}_i \big(\tilde{X}_i^\top G(x_i) \tilde{X}_i\big)^{-1} \tilde{X}_i^\top G(x_i),$$

where $(\cdot)_i$ is the i-th element of the vector argument.

4. Methods for Generating Prediction Sets

In this section, we discuss exact and approximate methods to generate prediction sets. These methods include sampling points directly from the prediction set boundary, as well as `rootCP` and `unionCP` approximation methods.

4.1. Sampling Boundary Points via Directional Root-Finding

Computing conformal prediction sets typically requires retraining the model for every candidate replacement of y_{n+1}, which becomes computationally infeasible in high dimensions. To address this, we introduce an efficient approach for multi-output settings by approximating the boundary of the conformal prediction set through directional root-finding. Our method fixes a finite set of directions $d \in \mathbb{R}^q$ and samples boundary points along each line $z(t, d) = z_0 + td$, where z_0 is a feasible interior point. These roots solve:

$$S_{n+1}(z(t^\star, d)) = S_i(z(t^\star, d)),$$

where S_i and S_{n+1} are conformity scores of observed and test points, respectively. Assuming linear predictors $\hat{y}(z) = Hy(z)$, the conformity scores become:

$$S_i(z(t,d)) = \|a_i - tb_i\|, \quad S_{n+1}(z(t,d)) = \|a_{n+1} - tb_{n+1}\|,$$

with

$$a_i = y_i - (Hy(z_0))_i, \quad a_{n+1} = z_0 - (Hy(z_0))_{n+1}, \quad b_i = H_{n+1}d, \quad b_{n+1} = d - H_{n+1}d.$$

This reduces the conformal set intersection to solving a 1D root-finding problem

$$\psi(t) = S_{n+1}(z(t,d)) - S_i(z(t,d)) = 0.$$

Solving with ℓ_1 Score For ℓ_1 norms, we express:

$$\psi(t) = \|a_{n+1} - tb_{n+1}\|_1 - \|a_i - tb_i\|_1 = c(t) + ts(t),$$

where both $c(t)$ and $s(t)$ are piecewise constant, changing sign at:

$$a_{i,j} - tb_{i,j} = 0 \quad \text{or} \quad a_{n+1,j} - tb_{n+1,j} = 0.$$

This yields up to $2q$ change-points $\{t_1^\star, \ldots, t_{2q}^\star\}$, from which roots are computed via:

$$\hat{t}_k = -\frac{c(t_k^\star)}{s(t_k^\star)}.$$

Solving with Mahalanobis Score For general Mahalanobis scores, squaring the norm simplifies computation:

$$\psi(t) = \|a_{n+1} - tb_{n+1}\|_M^2 - \|a_i - tb_i\|_M^2 = At^2 + Bt + C,$$

Roots are then obtained via the quadratic formula with coefficients:

$$A = \|b_{n+1}\|_M^2 - \|b_i\|_M^2, \quad B = -2(\langle a_{n+1}, b_{n+1}\rangle_M - \langle a_i, b_i\rangle_M), \quad C = \|a_{n+1}\|_M^2 - \|a_i\|_M^2.$$

Explicit Conformal Set Construction

Let $\{t_1, \ldots, t_K\}$ be the sorted intersection times. For any scalar t, we define:

$$(n+1)\pi(z(t,d)) = \sum_{i=1}^{n+1} \mathbb{1}_{t \in \mathcal{E}_i},$$

where $\mathcal{E}_i = \{z : S_{n+1}(z) \leq S_i(z)\}$ along the line $z_0 + td$. The conformal region along direction d is then:

$$\Gamma^{(\alpha)}(x_{n+1}, d) = \bigcup_{j:N(j)>(n+1)\alpha} (t_j, t_{j+1}) \cup \bigcup_{j:M(j)>(n+1)\alpha} \{t_j\},$$

where $N(j)$ counts how many intervals (t_j, t_{j+1}) lie in $\mathcal{E}_i$, and $M(j)$ counts how many t_j endpoints are included. This slicing approach reduces multidimensional conformal set approximation to tractable 1D subproblems, enabling efficient boundary approximation using convex sets such as ellipses or hulls.

4.2. Root-Based Approach Approximation via `rootCP`

Current efficient approaches to exact computation, limited to dimension one, are restricted to models that are piecewise-linear; this structure allows to track changes in the conformity function. We have extended these approaches to higher dimensions in the previous section.

To go beyond linear structures, we can use approximate homotopy approaches as in Ndiaye and Takeuchi (2019) which, given an optimization tolerance, provided a discretization of all the values that y_{n+1} can take. However, these approaches are also limited in dimension one and have an exponential complexity in the dimension of y_{n+1}. Convexity assumptions are also required, which, unfortunately, are not verified for more complex prediction models.

We extend the approximations of conformal prediction in multiple dimensions by computing conformal prediction set boundaries directly. Unlike the one-dimensional case where the boundary is often two points, in multiple dimensions the boundary is continuous and, thus, uncountable, which makes finite-time computation impossible. To get around this difficulty, we will first fix a finite set of search directions; we will estimate the intersection points between the boundary of the conformal prediction set and the chosen direction. Then, we use the points on the boundary as a data base to fit a convex approximation, *e.g.,* an ellipse or the convex hull, passing through these points. This estimates the set described in Equation 7 and is formally described below.

We suppose that the conformal prediction set is *star-shaped*, *i.e.,* there exists a point z_0 such that any other point z within $\Gamma^{(\alpha)}(x_{n+1})$ can be connected to z_0 with a line segment within the set. We note that a star-shaped set are note necessarily convex; ellipsoidal sets (or any convex set) are inherently star-shaped.

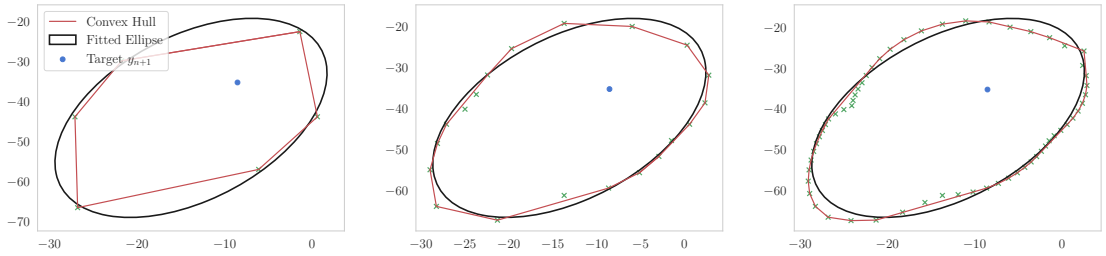


Figure 2: Illustration of the approximated conformal prediction set obtained fitting ellipse and convex hull given boundary points obtained by `rootCP` with various directions: three (left), ten (middle), and thirty (right). We use scikit-learn `make_regression` to generate synthetic dataset with the parameters n_samples $= 15$, n_features $= 5$, n_targets $= 2$ is the dimension of in output y_{n+1}.

Outline of `rootCP` For a direction $d \in \mathbb{R}^q$, the intersection points between the boundary of $\Gamma^{(\alpha)}(x_{n+1})$ and the line passing through z_0 and directed by d are obtained by solving the one dimensional equation

$$\pi(z(t,d)) = 1 - \alpha, \text{ where } z(t,d) = z_0 + td. \tag{19}$$

We briefly describe the main steps:

1. Fit a model μ_0 on the observed training set $\mathcal{D}_n$ and predict a feasible point $z_0 = \mu_0(x_{n+1})$.

2. For a collection of search directions $\{d_1, \ldots, d_K\}$, perform a bisection search in $[t_{\min}, 0]$ and $[0, t_{\max}]$ to output solutions $\hat{\ell}(d_k)$ and $\hat{u}(d_k)$ of Equation 19 at direction d_k, after at most $\log_2(\frac{t_{\max} - t_{\min}}{\epsilon_r})$ iterations for an optimization tolerance $\epsilon_r > 0$. Notice that the star-shape assumption implies that we will have only two roots on the selected directions.

3. Fit a convex set on the roots obtained at the previously $\{\hat{\ell}(d_k), \hat{u}(d_k)\}_{k \in [K]}$. In practice, when one uses a least-squares ellipse as a convex approximation, a number of search directions K proportional to the dimension q of the target y_{n+1} is sufficient. This is not necessarily the case for the convex hull. We refer to Figure 2 where we observe that many more search directions are needed to cover the conformal set when using the convex hull approximation. We also note that a minimum volume ellipsoid generated from the solutions generated would also be an appropriate approach to deliver a conservative convex approximation to the true prediction set.

4.3. `unionCP` Approximation Method

After constructing the set $\mathcal{E}$ for an incoming point x_{n+1}, it is initially unclear which regions $\mathcal{E}_i$ make up various conformal prediction sets, let alone how we need to combine these regions to get the exact conformal prediction sets. Thus, we aim to provide an approximation of conformal prediction sets using the regions generated with the approaches introduced in

Section 3. We provide Proposition 4 to bound error probabilities associated with potential combinations of these regions.

Proposition 4 *Let y_{n+1} be such that $(x_1, y_1), \ldots, (x_{n+1}, y_{n+1})$ are drawn exchangeably from $\mathcal{P}$. Then, under the uniqueness of conformity measures, for a set $\mathcal{S} \subset [n]$, drawn exchangeably, it holds that*

$$\mathbb{P}\Big(y_{n+1} \in \bigcup_{i \in \mathcal{S}} \mathcal{E}_i\Big) \geq \frac{|\mathcal{S}|}{n+1}.$$

Proof

First, consider a set $\mathcal{S} \subseteq [n]$. For any $z \notin \bigcup_{i \in \mathcal{S}} \mathcal{E}_i$

$$\begin{aligned} z \notin \bigcup_{i \in \mathcal{S}} \mathcal{E}_i \Leftrightarrow S_i(z) \leq S_{n+1}(z) \; \forall \; i \in \mathcal{S} \\ \Rightarrow \sum_{i=1}^{n+1} \mathbb{1}_{S_i(z) \leq S_{n+1}(z)} \geq |\mathcal{S}| + 1. \end{aligned}$$

Then, using a set $\mathcal{S}$ drawn exchangeably from $[n]$, for y_{n+1}

$$\begin{aligned} \mathbb{P}\Big(y_{n+1} \notin \bigcup_{i \in \mathcal{S}} \mathcal{E}_i\Big) &\leq \mathbb{P}\Big(\sum_{i=1}^{n+1} \mathbb{1}_{S_i(y_{n+1}) \leq S_{n+1}(y_{n+1})} \geq |\mathcal{S}| + 1\Big) \\ \Rightarrow \mathbb{P}\Big(y_{n+1} \in \bigcup_{i \in \mathcal{S}} \mathcal{E}_i\Big) &\geq 1 - \mathbb{P}\Big(\sum_{i=1}^{n+1} \mathbb{1}_{S_i(y_{n+1}) \leq S_{n+1}(y_{n+1})} \geq |\mathcal{S}| + 1\Big) \\ &\geq \mathbb{P}\Big(\sum_{i=1}^{n+1} \mathbb{1}_{S_i(y_{n+1}) \leq S_{n+1}(y_{n+1})} \leq |\mathcal{S}|\Big). \end{aligned}$$

Then, by Lemma 1, with the selection of $\alpha = 1 - \frac{|\mathcal{S}|}{n+1}$,

$$\mathbb{P}\Big(\sum_{i=1}^{n+1} \mathbb{1}_{S_i(y_{n+1}) \leq S_{n+1}(y_{n+1})} \leq |\mathcal{S}|\Big) \geq \frac{|\mathcal{S}|}{n+1}.$$

■

Proposition 4 states that with the exchangeable selection of a subset of $\mathcal{E}$, the probability of the response y_{n+1} being contained in the union of that subset is bounded-below by a function of cardinality. For example, if we wish to construct, say, a conservative 50% prediction set, we could select (exchangeably) a set $\mathcal{S} \subset [n]$ such that $|\mathcal{S}| \geq |\mathcal{E}|/2$; the union of all sets within $\mathcal{S}$ would provide a conservative prediction set, on average. We note that while our work emphasizes the use of $||\cdot||_W^2$, Proposition 4 holds for any conformity measure *e.g.,* ℓ_1.

While the union of a random selection of regions forms a conservative $1 - |\mathcal{S}|/(n+1)$ prediction set, we can provide more intelligently constructed sets that are empirically less conservative. Suppose we provide an ordering of our regions, where $\mathcal{E}_{(k)}$ is defined as the k-th smallest region by volume.

Definition 5 (`unionCP`) *A smaller* $(1-\alpha)$ *prediction set approximation can then be constructed as*

$$\hat{\Gamma}^{(\alpha)}(x_{n+1}) = \bigcup_{i \in \mathcal{S}_{1-\alpha}} \mathcal{E}_{(i)}, \tag{20}$$

where $\mathcal{S}_{1-\alpha} = [\lceil (1-\alpha)(n+1) \rceil]$. *We dub the approximation shown in Equation 20 as* `unionCP`.

By Proposition 4, `unionCP` generates an approximation that, at minimum, provides a region that is at least valid. We note that the set $\mathcal{S}_{1-\alpha}$ only depends on the the data values, and is, thus, generated exchangeably.

We compare prediction sets constructed using `unionCP` to a random selection of regions for multiple predictors in Section 5. We find empirically that sets constructed using `unionCP` are less conservative than a random collection of p-value change-point sets. Another approach using the union of regions generated through conformal prediction was introduced in Patel et al. (2024). We include an example of a prediction set approximation constructed with `unionCP` in Figure 4.

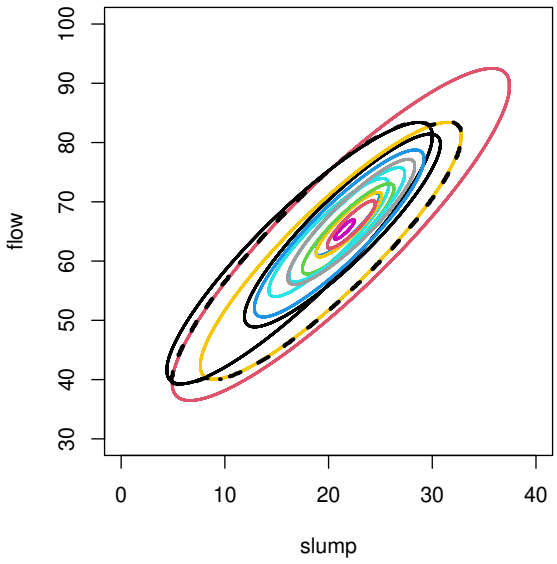


Figure 3: Example of `fullCP` prediction set (dotted black line) compared to the `unionCP` prediction set (union of colored p-value change-point sets) constructed using $||\cdot||_W^2$ with $W = \hat{\Sigma}^{-1}$ for an observation from `cement` dataset.

While Proposition 4 and the adjustment described in Equation 20 allow for conservative prediction sets, at times, the union of various $\mathcal{E}_i$ does not explicitly describe a conformal prediction set exactly. Thus, `unionCP` provides (at worse) a conservative approximation of the true conformal prediction set. We also note that `unionCP` requires the computation of volumes for each change-point set. While this is not an issue when using ℓ_1 or $||\cdot||_W^2$ as the conformity measure, it might prohibitive for other conformity measures.

5. Empirical Results

To provide empirical support for `unionCP` and `rootCP`, we consider three multi-output regression data sets, shown in Table 2.

Table 2: Summary of data sets used in empirical exploration.

Data	# features	# responses	Source
enb	8	2	Tsanas and Xifara (2012)
jura	13	3	Goovaerts (1997)
cement	7	3	Yeh (2007)

We first include results related to the empirical coverage of `unionCP` sets for our data sets of interest. We perform a small exploration, comparing the coverage of a random selection of p-value change-point sets generated with linear regression (LS) to `unionCP` with other prediction methods. The results are shown in Figure 4. We note that the randomized approach significantly overcovers, while `unionCP` provides well-calibrated prediction sets.

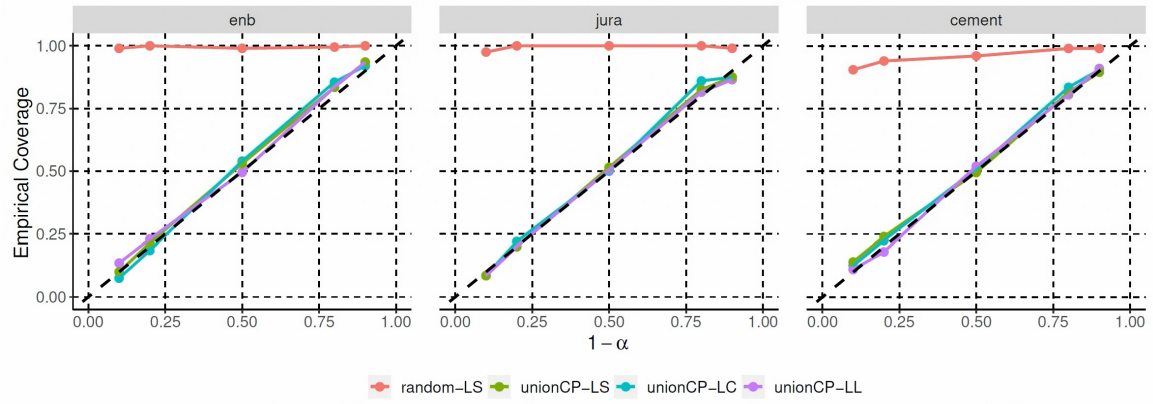
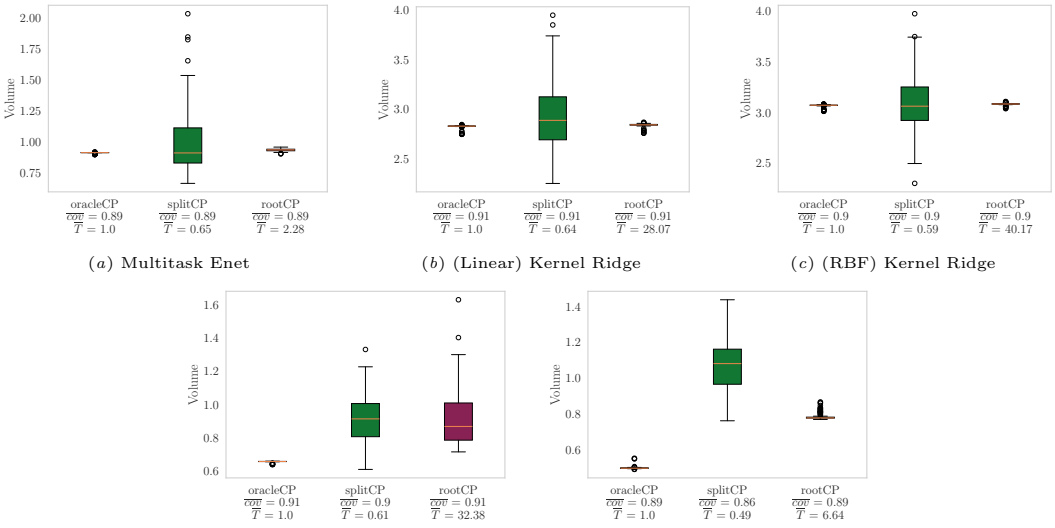


Figure 4: Comparison of empirical coverage with random selection of k regions and `unionCP` for various datasets and predictors, including: linear regression (LS), Nadaraya-Watson (LC), and local-linear (LL) across 100 repetitions with $n = 40$.

For our `rootCP` comparision, we consider a reference benchmark; specifically, we have $\pi(y_{n+1}) \geq \alpha$ with probability larger than $1 - \alpha$. Hence, we can define the `oracleCP` as $\pi^{-1}([\alpha, +\infty))$ where π is obtained with a model fit optimized on the oracle data $\mathcal{D}_{n+1}(y_{n+1})$ on top of the root-based approach to find boundary points. We remind the reader that the target variable y_{n+1} is not available in practice. The experiments were conducted with a coverage level of 0.9, *i.e.,* $\alpha = 0.1$. We include results for a suite of complex regression models in Figures 5-7.

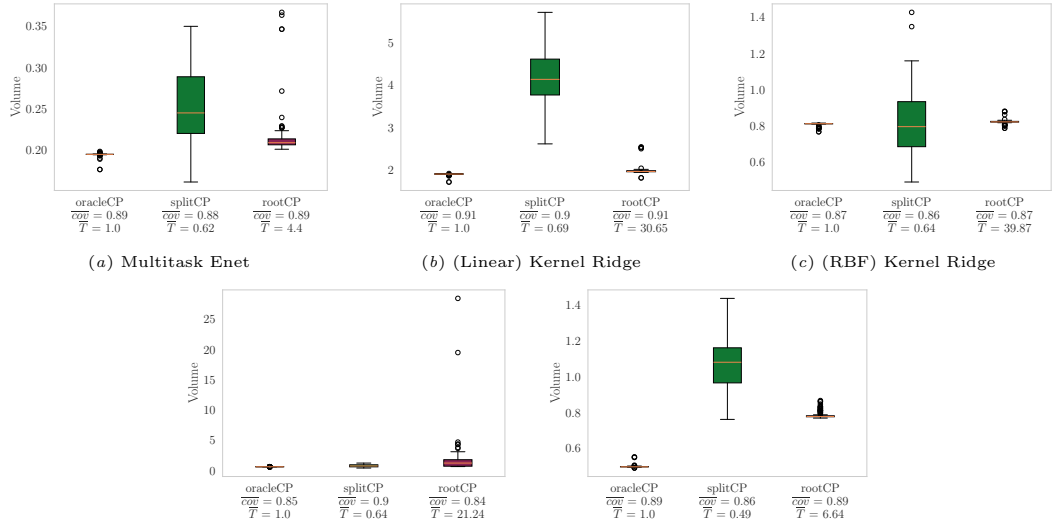
We note that in all cases across all datasets, the `rootCP` approach is competitive with `splitCP`, and in most cases, surpasses `splitCP` in terms of volume. This result provides a trade-off between generating low-volume sets (with high computational load) through `fullCP` and high-volume sets through `splitCP` (with low computational load).



(*a*) Multitask Enet (*b*) (Linear) Kernel Ridge (*c*) (RBF) Kernel Ridge

(*d*) Support Vector Regression (*e*) *k*-NN

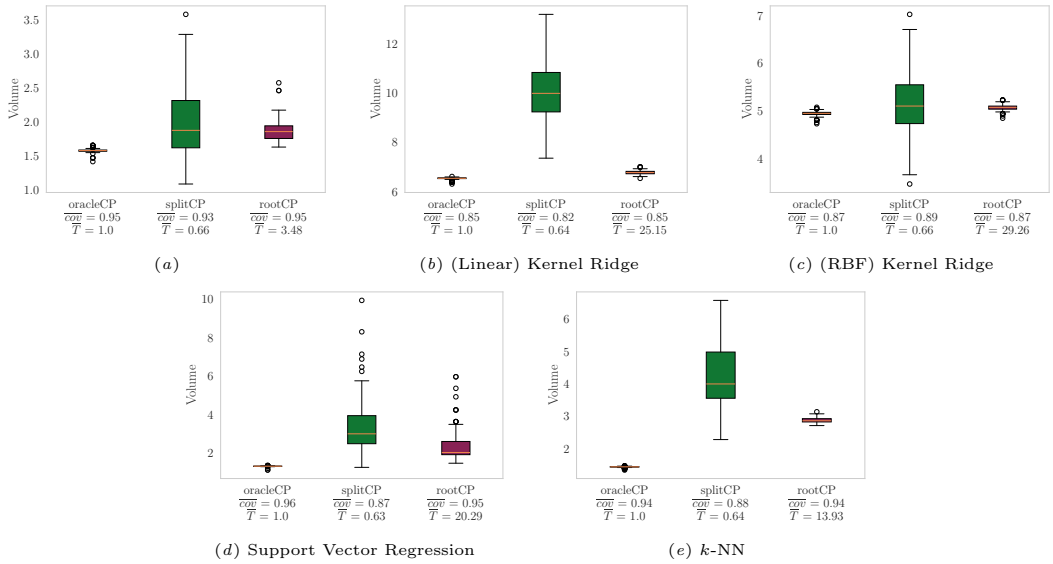
Figure 5: Benchmarking the ellipse based conformal sets for several regression models on `enb` dataset. We display the lengths of the confidence sets over 100 random permutation of the data. We denoted $\overline{cov}$ the average coverage, and $\overline{T}$ the average computational time normalized with the average time for computing `oracleCP` which requires a single model fit on the whole data.



(*a*) Multitask Enet (*b*) (Linear) Kernel Ridge (*c*) (RBF) Kernel Ridge

(*d*) Support Vector Regression (*e*) *k*-NN

Figure 6: Benchmarking the ellipse based conformal sets for several regression models on `jura` dataset. We display the lengths of the confidence sets over 100 random permutation of the data. We denoted $\overline{cov}$ the average coverage, and $\overline{T}$ the average computational time normalized with the average time for computing `oracleCP` which requires a single model fit on the whole data.



(*a*) (*b*) (Linear) Kernel Ridge (*c*) (RBF) Kernel Ridge

(*d*) Support Vector Regression (*e*) *k*-NN

Figure 7: Benchmarking the ellipse based conformal sets for several regression models on `cement` dataset. We display the lengths of the confidence sets over 100 random permutation of the data. We denoted $\overline{cov}$ the average coverage, and $\overline{T}$ the average computational time normalized with the average time for computing `oracleCP` which requires a single model fit on the whole data.

6. Conclusion

In this paper, we introduced exact p-values in multiple dimensions for predictors that are a linear function of the candidate value. Specifically, we discussed the exact construction of p-values using $||\cdot||_W^2$. Additionally, we introduced a multi-output extension to `rootCP`, which can be used with a wide class of predictors. We also introduced `unionCP`, which allows for approximation of the conformal prediction set for linear predictors. We then showed empirical results with multiple predictors, including a subset of both linear and nonlinear predictors, while drastically reducing the computational requirements compared to `fullCP`.

Other questions about the theoretical guarantees of our approach have yet to be answered. For example, we lack precise characterizations on the number of points to be sampled on the conformal set boundary. Besides the conformal sets presented in this paper, these questions are equally relevant to the construction of any high-dimensional confidence sets. Future work might also explore the performance of these approximations with a higher dimension response and under conditions related to dependence or covariate shift.

References

Jonathan Alvarsson, Staffan Arvidsson McShane, Ulf Norinder, and Ola Spjuth. Predicting with confidence: using conformal prediction in drug discovery. *Journal of Pharmaceutical Sciences*, 110(1):42–49, 2021.

Anastasios N Angelopoulos, Stephen Bates, et al. Conformal prediction: A gentle introduction. *Foundations and Trends® in Machine Learning*, 16(4):494–591, 2023.

Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. Predictive inference with the jackknife+. *The Annals of Statistics*, 49(1):486–507, 2021.

Umang Bhatt, Adrian Weller, and Giovanni Cherubin. Fast conformal classification using influence functions. In *Conformal and Probabilistic Prediction and Applications*, pages 303–305. PMLR, 2021.

Hanen Borchani, Gherardo Varando, Concha Bielza, and Pedro Larranaga. A survey on multi-output regression. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(5):216–233, 2015.

Lars Carlsson, Martin Eklund, and Ulf Norinder. Aggregated conformal prediction. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pages 231–240. Springer, 2014.

Leonardo Cella and Ryan Martin. Valid distribution-free inferential models for prediction. *arXiv preprint arXiv:2001.09225*, 2020.

Wenyu Chen, Zhaokai Wang, Wooseok Ha, and Rina Foygel Barber. Trimmed conformal prediction for high-dimensional models. *arXiv preprint arXiv:1611.09933*, 2016.

Giovanni Cherubin, Konstantinos Chatzikokolakis, and Martin Jaggi. Exact optimization of conformal predictors via incremental and decremental learning. In *International Conference on Machine Learning*, pages 1836–1845. PMLR, 2021.

Alex Contarino, Christine Shubert Kabban, Chancellor Johnstone, and Fairul Mohd-Zaid. Constructing prediction intervals with neural networks: an empirical evaluation of bootstrapping and conformal inference methods. *arXiv preprint arXiv:2210.05354*, 2022.

Isidro Cortés-Ciriano and Andreas Bender. Concepts and applications of conformal prediction in computational drug discovery. *arXiv preprint arXiv:1908.03569*, 2019.

Jacopo Diquigiovanni, Matteo Fontana, and Simone Vantini. Conformal prediction bands for multivariate functional data. *Journal of Multivariate Analysis*, 189:104879, 2022.

Martin Eklund, Ulf Norinder, Scott Boyer, and Lars Carlsson. The application of conformal prediction to the drug discovery process. *Annals of Mathematics and Artificial Intelligence*, 74:117–132, 2015.

Jianqing Fan. Design-adaptive nonparametric regression. *Journal of the American statistical Association*, 87(420):998–1004, 1992.

Matteo Fontana, Gianluca Zeni, and Simone Vantini. Conformal prediction: a unified review of theory and new challenges. *Bernoulli*, 29(1):1–23, 2023.

Alexander Gammerman, Vladimir Vovk, and Vladimir Vapnik. Learning by transduction. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 148–155, 1998.

Pierre Goovaerts. *Geostatistics for natural resources evaluation*. Applied Geostatistics, 1997.

Marc Hallin, Davy Paindaveine, Miroslav Šiman, Ying Wei, Robert Serfling, Yijun Zuo, Linglong Kong, and Ivan Mizera. Multivariate quantiles and multiple-output regression quantiles: From l1 optimization to halfspace depth [with discussion and rejoinder]. *The Annals of Statistics*, pages 635–703, 2010.

Chancellor Johnstone and Bruce Cox. Conformal uncertainty sets for robust optimization. In *Conformal and Probabilistic Prediction and Applications*, pages 72–90. PMLR, 2021.

Arun Kumar Kuchibhotla. Exchangeability, conformal prediction, and rank tests. *arXiv preprint arXiv:2005.06095*, 2020.

Alexander Kuleshov, Alexander Bernstein, and Evgeny Burnaev. Conformal prediction in manifold learning. In *Conformal and Probabilistic Prediction and Applications*, pages 234–253, 2018.

Jing Lei. Fast exact conformalization of the lasso using piecewise linear homotopy. *Biometrika*, 106(4):749–764, 2019.

Jing Lei, Alessandro Rinaldo, and Larry Wasserman. A conformal prediction approach to explore functional data. *Annals of Mathematics and Artificial Intelligence*, 74(1):29–43, 2015.

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.

Soundouss Messoudi, Sébastien Destercke, and Sylvain Rousseau. Conformal multi-target regression using neural networks. In *Conformal and Probabilistic Prediction and Applications*, pages 65–83. PMLR, 2020.

Soundouss Messoudi, Sébastien Destercke, and Sylvain Rousseau. Copula-based conformal prediction for multi-target regression. *arXiv preprint arXiv:2101.12002*, 2021.

Soundouss Messoudi, Sébastien Destercke, and Sylvain Rousseau. Ellipsoidal conformal inference for multi-target regression. *Proceedings of Machine Learning Research*, 179:1–13, 2022.

Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9 (1):141–142, 1964.

Eugene Ndiaye. Stable conformal prediction sets. In *International Conference on Machine Learning*, pages 16462–16479. PMLR, 2022.

Eugene Ndiaye and Ichiro Takeuchi. Computing full conformal prediction set with approximate homotopy. *arXiv preprint arXiv:1909.09365*, 2019.

Eugene Ndiaye and Ichiro Takeuchi. Root-finding approaches for computing conformal prediction set. *arXiv preprint arXiv:2104.06648*, 2021.

Jelmer Neeven and Evgueni Smirnov. Conformal stacked weather forecasting. In *Conformal and Probabilistic Prediction and Applications*, pages 220–233, 2018.

Ilia Nouretdinov, Thomas Melluish, and Volodya Vovk. Ridge regression confidence machine. In *ICML*, pages 385–392. Citeseer, 2001.

Henrik Olsson, Kimmo Kartasalo, Nita Mulliqi, Marco Capuccini, Pekka Ruusuvuori, Hemamali Samaratunga, Brett Delahunt, Cecilia Lindskog, Emiel AM Janssen, Anders Blilie, et al. Estimating diagnostic uncertainty in artificial intelligence assisted pathology using conformal prediction. *Nature communications*, 13(1):7761, 2022.

Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

Yash P Patel, Sahana Rayan, and Ambuj Tewari. Conformal contextual robust optimization. In *International Conference on Artificial Intelligence and Statistics*, pages 2485–2493. PMLR, 2024.

David W Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 1992.

Ryan J Tibshirani, Rina Foygel Barber, Emmanuel Candes, and Aaditya Ramdas. Conformal prediction under covariate shift. *Advances in neural information processing systems*, 32, 2019.

Athanasios Tsanas and Angeliki Xifara. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and buildings*, 49:560–567, 2012.

Vladimir Vovk. Cross-conformal predictors. *Annals of Mathematics and Artificial Intelligence*, 74(1):9–28, 2015.

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*. Springer Science & Business Media, 2005.

Devin Wasilefsky, William Caballero, Chancellor Johnstone, Nathan Gaw, and Phillip Jenkins. Responsible machine learning for united states air force pilot candidate selection, 2023.

Geoffrey S Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372, 1964.

Donna Xu, Yaxin Shi, Ivor W Tsang, Yew-Soon Ong, Chen Gong, and Xiaobo Shen. Survey on multi-output learning. *IEEE transactions on neural networks and learning systems*, 31 (7):2409–2429, 2019.

I-Cheng Yeh. Modeling slump flow of concrete using second-order regressions and artificial neural networks. *Cement and concrete composites*, 29(6):474–480, 2007.

Yu Zhang and Qiang Yang. An overview of multi-task learning. *National Science Review*, 5 (1):30–43, 2018.