Electronic Journal of Statistics
Vol. 19 (2025) 6141–6166
ISSN: 1935-7524
https://doi.org/10.1214/25-EJS2478

Conformal inference for regression on Riemannian manifolds

Alejandro Cholaquidis[1], **Fabrice Gamboa**[2] **and Leonardo Moreno**[3]

[1]*Centro de Matemática, FCIEN, Universidad de la República, Uruguay,*
e-mail: acholaquidis@hotmail.com

[2]*Institut de Mathématiques de Toulouse, Université Toulouse, France,*
e-mail: fabrice.gamboa@math.univ-toulouse.fr

[3]*IESTA, DMMC, FCEA, Universidad de la República, Uruguay,*
e-mail: leonardo.moreno@fcea.edu.uy

Abstract: Regression on manifolds, and, more broadly, statistics on manifolds, have garnered significant importance in recent years due to the vast number of applications for non-Euclidean data. Circular data is a classic example, but so is data in the space of covariance matrices, data on the Grassmannian manifold obtained as a result of principal component analysis, among many others. In this work we investigate prediction sets for regression scenarios when the response variable, denoted by Y, resides in a manifold, and the covariate, denoted by X, lies in a Euclidean space. This extends the concepts delineated in Lei and Wasserman (2014) to this novel context. Aligning with traditional principles in conformal inference, these prediction sets are distribution-free, indicating that no specific assumptions are imposed on the joint distribution of (X, Y), and they maintain a non-parametric character. We prove the asymptotic almost-sure convergence of the empirical version of these regions on the manifold to their population counterparts. The efficiency of this method is shown through a comprehensive simulation study and an analysis involving real-world data.

Keywords and phrases: Confidence set, Grassmannian manifold, principal component analysis, wind intensity.

Received July 2025.

1. Introduction

Conformal prediction is a powerful set of statistical tools that operates under minimal assumptions about the underlying model. It is primarily used to construct confidence sets that are applicable to a diverse range of problems in fields such as machine learning and statistics. It is unique in its ability to construct prediction regions that guarantee a specified level of coverage for finite samples, irrespective of the distribution of the data. This is particularly crucial in high-stakes decision-making scenarios where maintaining a certain level of coverage is critical.

Unlike other methods, conformal prediction does not require strong assumptions about the sample distribution, such as normality. The aim is to construct prediction regions as small as possible to yield informative and precise predictions, enhancing the tool's utility in various applications.

The approach was first proposed by Vovk, Gammerman, and Shafer in the late 1990s, as referenced in Vovk et al. (1998). Since its inception, it has been the subject of intense research activity. Originally formulated for binary classification problems, the method has since been expanded to accommodate regression, multi-class classification, functional data, functional time series, and anomaly detection, among others. Several applications of this method can be found in the book Balasubramanian et al. (2014).

In the context of regression, conformal prediction has proven to be efficient in constructing prediction sets, as evidenced by works such as Lei and Wasserman (2014), Wäschle et al. (2014), and Kuleshov et al. (2018). To enhance the performance of these prediction sets, particularly to decrease the length of prediction intervals (when the output is one-dimensional), a combination of conformal inference and quantile regression was proposed in Romano et al. (2019). In Fong and Holmes (2021), computationally efficient conformal inference methods for Bayesian models were proposed.

The method has also been extended to functional regression, where the predictors and responses are functions, rather than vectors, see for instance Lei et al. (2015), Fontana et al. (2020) and Diquigiovanni et al. (2022). In this case, the prediction regions take the form of functional envelopes that have a high probability of containing the true function.

In the field of classification, conformal prediction has been employed to tackle a broad spectrum of problems. These include image classification, as in Lei (2014), and text classification, as in Vovk et al. (2023). For multi-class classification, a prevalent approach is to create prediction sets that have a high likelihood of encompassing the correct class. This can be realized via the use of 'Venn prediction sets', as outlined in Liu and Wasserman (2016). These sets partition the label space into overlapping regions, each one corresponding to a distinct class.

In summary, conformal prediction provides powerful statistical tools and has been successfully applied to a wide range of problems in machine learning, statistics, and related fields. Its main advantage is its ability to provide distribution-free prediction regions that can be used in the presence of any underlying distribution of the data. As this field of research continues to evolve, it is expected to find even more applications in the future.

2. Conformal inference on manifolds

Although conformal inference has been extended to many settings, no proposals have yet addressed the case where the response lies on a Riemannian manifold. This scenario is crucial for data that, by nature, belong to nonlinear spaces—such as covariance matrices on the manifold of positive definite matrices (Best, 2010; Huang and Kou, 2014; Cholaquidis et al., 2023), orthonormal frames on the Stiefel manifold (Cholaquidis et al., 2023), subspaces obtained from PCA on the Grassmannian (Hong et al., 2016), or directional measurements on cylindrical manifolds (Cholaquidis et al., 2022). Further applications appear in image

analysis (Pennec et al., 2019) and geometric machine learning (Guigui et al., 2023).

The extension to this context is not trivial since when the output Y belongs to a Riemannian manifold $\mathcal{M}$, several arguments used in Euclidean conformal inference are not straightforwardly extendable. For instance, the Fréchet mean on the manifold can be defined, but it does not always exist nor is it necessarily unique. Moreover, formulating a regression model on the manifold is not easy because of the absence of any additive structure, although there have been recent advances in this area (see for instance Petersen and Müller (2019)).

We extend the methodologies outlined by Lei and Wasserman (2014) to the broader context of pairs (X, Y) where $X \in \mathbb{R}^d$ and the response Y has support included on a, sufficiently smooth, ℓ-dimensional manifold $\mathcal{M}$. Lei and Wasserman (2014) is focused on the scenario where Y is a real-valued variable, constructing confidence sets—referred to as confidence bands—via density estimators. One of the key tools to get the consistency of the empirical conformal region to its population counterpart, as in Lei and Wasserman (2014), is Theorem 1. It states that the kernel-density estimator is uniformly consistent on manifolds. Once this is established, to adapt the ideas in Lei and Wasserman (2014), it must also be proved that the conditional density is uniformly bounded from above, this is done in Lemma 4. Finally, a further distinction from Lei and Wasserman (2014) is that—because the manifold may possess a nonempty boundary—points located well inside the manifold are handled differently from those situated near its edge. Our main and stronger theorem is for compact manifolds, however, we considered in Section 5 the case of non-compact manifolds.

Following Cholaquidis et al. (2022), the kernel estimator employs Euclidean distances, which may introduce bias in regions of high curvature or near the boundary. Section 6 explores computational approaches for high-dimensional or complex manifolds, inspired by Izbicki et al. (2022). Related work in general metric spaces (Lugosi and Matabuena, 2024) provides only consistency in probability, without explicit convergence rates as obtained here.

2.1. Conformal inference in a nutshell

In this section we briefly recall the foundations of conformal inference; see Fontana et al. (2023) for details. A key hypothesis is *exchangeability*: for any permutation π of $\{1, \ldots, n\}$, the distribution of $(Z_1, \ldots, Z_n)$ equals that of $(Z_{\pi(1)}, \ldots, Z_{\pi(n)})$, where each Z_i takes values in a measurable space $\mathbf{Z}$. Quoting Fontana et al. (2023): "A nonconformity measure $A(B, z) : \mathbf{Z}^n \times \mathbf{Z} \to \mathbb{R}$ scores how different an example z is from a bag $B = \{Z_1, \ldots, Z_n\}$. Define $p_z := |\{i = 1, \ldots, n+1 : R_i \geq R_{n+1}\}|/(n+1)$, where $R_i := A(\{Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_n, z\}, Z_i)$ and $R_{n+1} := A(\{Z_1, \ldots, Z_n\}, z)$." For $\alpha \in [0, 1]$, the prediction set is

$$\gamma^{\alpha}(Z_1, \ldots, Z_n) := \{z \in \mathbf{Z} : p_z > \alpha\}.$$

The following proposition (Proposition 2.1 in (Vovk et al., 2023)) holds.

Proposition 1. *Under exchangeability,*

$$\mathbb{P}\big(Z_{n+1} \notin \gamma^{\alpha}(Z_1, \ldots, Z_n)\big) \leq \alpha \quad \textit{for any } \alpha \in [0,1].$$

In the regression setting, that is, when $Z_i = (X_i, Y_i)$ has distribution P, previous construction builds a level set on the joint distribution (see Equation 4 in Lei and Wasserman (2014)). Nevertheless, controlling $\mathbb{P}(Y \in \mathbf{C}_P(x)|X = x)$ is much more convenient in regression (see Lei and Wasserman (2014)). In this case we speak of "conditional coverage". When there exists a conditional density $p(y|x)$ (see hypothesis H1 in subsection 3), the "conditional oracle set", $\mathbf{C}_P(x)$, (in Lei and Wasserman (2014) it is called conditional oracle band), is defined as

$$\mathbf{C}_P(x) = \{y : p(y|x) \geq t_x^{\alpha}\}, \tag{2.1}$$

where $t^{\alpha}(x)$ satisfies

$$\int \mathbf{1}_{\{p(y|x) \geq t_x^{\alpha}\}} p(y|x) dy = 1 - \alpha. \tag{2.2}$$

The empirical version of $\mathbf{C}_P(x)$ obtained from $\aleph_n = \{(X_i, Y_i) : i = 1, \ldots, n\}$ constituted by an i.i.d. sample of (X, Y) with distribution P, is called a conditionally valid set $\mathbf{C}_n(x)$ (see Lei and Wasserman (2014)).

Definition 1. Given $\alpha \in (0,1)$, and x in the support of P_X, a set $\mathbf{C}_n(x) \subset \mathcal{M}$ is said to be **conditionally valid** if

$$\mathbb{P}\big(Y \in \mathbf{C}_n(x)|X = x\big) \geq 1 - \alpha.$$

This definition captures the notion of a set of possible values of Y that provides a specified level of coverage for a given input x.

As is shown in Lemma 1 of Lei and Wasserman (2014), non-trivial finite-sample conditional validity for all x in the support of P_X is impossible for a continuous distribution. To overcome this limitation, the following notion of local validity is introduced.

Definition 2. Let $\mathcal{A} = \{A_j : j \geq 1\}$ be a partition of $\text{supp}(P_X)$. A prediction set $\mathbf{C}_n$ is **locally valid** with respect to $\mathcal{A}$ if

$$\mathbb{P}(Y_{n+1} \in \mathbf{C}_n(X_{n+1})|X_{n+1} \in A_j) \geq 1 - \alpha \text{ for all } j \text{ and all } P. \tag{2.3}$$

Whenever $A_k \in \mathcal{A}$, we will write $p(y|A_k)$ for the conditional density, w.r.t. ν, of Y given $X \in A_k$. We aim to prove (see Theorem 2) that locally valid sets converges to $\mathbf{C}_P$ when Y is supported in a manifold.

3. Assumptions

In the following, $\aleph_n = \{(X_i, Y_i) : i = 1, \ldots, n\}$ denotes an i.i.d. sample of (X, Y) with distribution P, where $X \in \mathbb{R}^d$ and $Y \in \mathcal{M}$. Here $(\mathcal{M}, \rho)$ is a compact ℓ-dimensional submanifold of $\mathbb{R}^D$, whose geodesic distance is denoted by ρ. The

case of non-compact manifolds is discussed in Section 5. We further denote by P_X the marginal distribution of X and by $\text{supp}(P_X)$ its support. We denote the volume measure on $\mathcal{M}$ by ν and use $\|\cdot\|$ to denote the Euclidean norm on $\mathbb{R}^D$. We denote by μ the d-dimensional Lebesgue measure in $\mathbb{R}^d$.

We will now give the set of assumptions that we will require. H1 to H4 are also imposed in Lei and Wasserman (2014). Hypotheses H0 and H5 are imposed to guarantee the uniform convergence of the kernel-based density estimator of the conditional density.

- H0 $\mathcal{M}$ is a $\mathcal{C}^2$ submanifold, and if $\partial\mathcal{M} \neq \emptyset$, then $\partial\mathcal{M}$ is also a $\mathcal{C}^2$ submanifold.
- H1 The joint distribution P has a density $p_{X,Y}(x,y)$ w.r.t. $\mu \times \nu$, and the marginal distribution P_X has a density p_X (w.r.t. μ). We denote by $p(y|x) = p_{X,Y}(x,y)/p_X(x)$ the conditional density, where $p(y|x) = 0$ if $p_X(x) = 0$.
- H2 p_X is such that there exist b_1, b_2 such that $0 < b_1 \leq p_X(x) \leq b_2 < \infty$, for all $x \in \text{supp}(P_X)$
- H3 $p(y|x)$ is Lipschitz continuous as a function of x, i.e., there exists a constant $L > 0$ such that $\|p(\cdot|x) - p(\cdot|x')\|_\infty \leq L\|x - x'\|$.
- H4 There are positive constants ϵ_0, γ, c_1, and c_2 such that for all x in the support of X,

$$c_1\epsilon^\gamma \leq \nu\big(\{y : |p(y|x) - t_x^\alpha| < \epsilon\}\big) \leq c_2\epsilon^\gamma.$$

  for all $\epsilon \leq \epsilon_0$, where t_x^α is given by (2.2). Moreover $\inf_{x\in\text{supp}(P_X)} t_x^\alpha \geq t_0 > 0$.
- H5 For all k, $p(\cdot|A_k) \in \mathcal{C}^2$.

H0 is satisfied by a wide range of manifolds, among them the Stiefel and Grassmannian manifolds, and the cone of positive-definite matrices; it is not restrictive in applications. Although the $\mathcal{C}^2$ regularity is sufficient for our theoretical results, in practical kernel-based estimation the local curvature and the injectivity radius of the manifold $\mathcal{M}$ may influence performance. High-curvature regions or small injectivity radius reduce the domain where geodesic distances behave approximately Euclidean, which can affect bias when Euclidean distances are used in the kernel. In such cases, smaller bandwidths or curvature-adaptive kernels (see, e.g., Berry and Sauer (2017)) may help mitigate distortion near regions of high curvature or close to the cut locus.

Assumptions H1 to H5 impose regularity on the joint and conditional densities of the data. In particular, the assumption of Lipschitz continuity H3 ensures that small changes in the input x lead to small changes in the output y. Assumption H2 implies that X is compactly supported. Regarding Assumption H4, quoting Lei and Wasserman (2014), "is related to the notion of the 'γ-exponent' condition that was introduced by Polonik (1995), and widely used in the density level set literature (Tsybakov (1997); Rigollet and Vert (2009)). It ensures that the conditional density $p(\cdot|x)$ is neither too flat nor too steep near the contour at level t_x^α, so the cut-off value t_x^α and the conditional density level set $\mathbf{C}_P(x) = L_x(t_x^\alpha)$ can be approximated from a finite-sample [...]. Assumption H4 also requires that the optimal cut-off values t_x^α be bounded away from zero.".

Assumption H5 is required to get the uniform convergence of the Kernel based estimator of $p(y|A_k)$ on manifolds, see Berry and Sauer (2017). These assumptions play a crucial role in the development and analysis of conformal prediction methods.

Let us introduce some important sequences of positive real numbers that will play a key role all along the manuscript.

1. $w_n = (\log(n)/n)^{1/(d+2)}$,
2. $\gamma_n = \lfloor b_1 n w_n^d/2 \rfloor$, being b_1 the positive constant introduced in H2. Observe that $\gamma_n \to \infty$.
3. Given $h_n, c_n \to 0$ as $n \to \infty$, we will consider the subsequences c_{γ_n} and h_{γ_n} being $\gamma_n = \lfloor b_1 n w_n^d/2 \rfloor$ as before. We assume h_n is strictly decreasing.

4. Locally valid sets from a kernel density estimator

In this section we introduce a slightly modified version of the estimated local marginal density $\hat{p}^{(x,y)}(v|A_k)$ and of the local conformity rank $\pi_{n,k}(x,y)$ originally introduced in Lei and Wasserman (2014). We make the assumption that $\text{supp}(P_X) = [0,1]^d$, and that $\mathcal{A} = \{A_k, k = 1, \ldots, T\}$ is a finite partition of $[0,1]^d$ consisting of equilateral cubes with sides of length w_n. This is a common technical assumption in conformal inference, but we can assume that, for instance, X is such that $\text{supp}(P_X) \subset [-R,R]^d$, for some $R > 0$. It only changes the constants appearing in Theorem 2, which still holds.

Let $n_k = \sum_{i=1}^n \mathbf{1}_{\{X_i \in A_k\}}$. Given a sequence $h_n \to 0$, a kernel function $K(\cdot) : \mathbb{R} \to \mathbb{R}$, and h_{n_k}, we define

$$\hat{p}(y|A_k) = \frac{1}{n_k h_{n_k}^\ell} \sum_{i=1}^n \mathbf{1}_{\{X_i \in A_k\}} K\left(\frac{\|Y_i - y\|}{h_{n_k}}\right). \tag{4.4}$$

We aim to prove that $\hat{p}(y|A_k)$ provides a uniform estimate of $p(y|A_k)$ across y and k. To achieve this, we need to ensure that there are sufficiently many sample points in each A_k. This is guaranteed by lemma 9 of Lei and Wasserman (2014), which states that if we choose $w_n = (\log(n)/n)^{1/(d+2)}$, then, with probability one, for all n large enough,

$$\forall k: \quad b_1 n w_n^d/2 \le n_k \le 3 b_2 n w_n^d/2. \tag{4.5}$$

Recall that b_1 and b_2 were defined in Assumption H2. In what follows we work on the event $\mathcal{E}_n = \{\forall k: \ b_1 n w_n^d/2 \le n_k \le 3 b_2 n w_n^d/2\}$, which holds a.s. for all large n.

The corresponding augmented estimate, based on $\aleph_n \cup \{(x,y)\}$ is, for any $(x,y) \in A_k \times \mathcal{M}$,

$$\hat{p}^{(x,y)}(v \mid A_k) = \frac{n_k}{n_k + 1} \hat{p}(v \mid A_k) + \frac{1}{(n_k+1)\, h_{n_k}^\ell} K\left(\frac{\|y - v\|}{h_{n_k}}\right).$$

For any $(X_{n+1}, Y_{n+1}) = (x, y) \in A_k \times \mathcal{M}$, consider the following **local conformity rank**

$$\pi_{n,k}(x,y) = \frac{1}{n_k+1} \sum_{i=1}^{n+1} \mathbf{1}_{\{X_i \in A_k\}} \mathbf{1}_{\{\hat{p}^{(x,y)}(Y_i|A_k) \leq \hat{p}^{(x,y)}(Y_{n+1}|A_k)\}}. \tag{4.6}$$

As proved in Proposition 2 of Lei and Wasserman (2014), the set

$$\hat{\mathbf{C}}_n(x) = \{y : \pi_{n,k}(x,y) \geq \alpha\}, \tag{4.7}$$

has finite-sample local validity, i.e., it satisfies (2.3). Finite-sample local validity, as established by (4.7), does not require any assumptions and it holds under very general conditions.

The following result is the key theorem. Its proof is deferred to the Appendix A. The proof follows the ideas used to prove Theorem 1 of Cholaquidis et al. (2022). It states that $p(y|A_k)$ can be estimated uniformly by (4.4) for all y that are far enough from the boundary of $\mathcal{M}$. Additionally, this estimation can be made uniformly across all k. We assume, as in Cholaquidis et al. (2022), that K is a Gaussian kernel. This restriction is, as in Cholaquidis et al. (2022), purely technical. More recent references on kernel-density estimation on manifolds are Bouzebda and Taachouche (2024a,b), Cleanthous et al. (2020, 2022); Berenfeld and Hoffmann (2021); Wu and Wu (2022).

This result is of fundamental importance in conformal prediction, as it provides a way to estimate the conditional density of Y given A_k for any k in a nonparametric way. It implies that the estimation error is uniform across all partitions A_k, which is a key requirement for conformal prediction methods. In particular, it allows us to construct conformal prediction regions that are valid with a given level of confidence.

Theorem 1. *Let $\mathcal{M} \subset \mathbb{R}^D$ be a compact ℓ-dimensional manifold satisfying H0. Let $\mathcal{M}_n \subset \mathcal{M}$ be a sequence of closed sets, and let $h_n \to 0$ be a sequence of bandwidths such that $nh_n^{\ell+3}/\log(n) \to \infty$. We assume that $c_n := \inf_{y \in \mathcal{M}_n} \rho(y, \partial\mathcal{M}) \downarrow 0$ is such that $h_n/c_n \downarrow 0$ monotonically. Additionally, we assume that $p(y|A_k)$ satisfies H5. Then, we have*

$$\sup_k \sup_{y \in \mathcal{M}_{n_k}} |\hat{p}(y|A_k) - p(y|A_k)| = o(h_{\gamma_n}/c_{\gamma_n}) \quad a.s., \tag{4.8}$$

where $\gamma_n = \lfloor b_1 n w_n^d/2 \rfloor$, and $n_k = \sum_{i=1}^n \mathbf{1}_{\{X_i \in A_k\}}$.

The following theorem is the main result of our paper. It states that $\hat{\mathbf{C}}_n(X_i)$ consistently estimates $\mathbf{C}_P(X_i)$ uniformly for all $X_i \in \aleph_n$. The proof, which is deferred to the Appendix, see subsection A.1, is based on some ideas from the proof of Theorem 1 of Lei and Wasserman (2014). Specifically, Lemma 1 and Lemma 3 are adapted from Lei and Wasserman (2014), while Lemma 2 is new and is required to adapt the proof of Lemma 3. Lemma 4 is the same as Lemma 8 of Lei and Wasserman (2014). Finally, the proof of Theorem 2 uses the adapted lemmas and considers separately the sample points that are close to $\mathcal{M}$ and

those that are far away from this boundary. The first set of points is shown to be negligible with respect to ν using techniques from geometric measure theory.

Theorem 2. *Assume H0 to H5. Let* $\hat{\mathbf{C}}_n(x)$ *be given by* (4.7) *and* $\mathbf{C}_P(x)$ *be given by* (2.1). *Let* $w_n = (\log(n)/n)^{1/(d+2)}$ *and* $\gamma_n = \lfloor b_1 n w_n^d/2 \rfloor$. *Then, for any* $\lambda > 0$, *there exists an* $A_\lambda > 0$ *such that, for* n *large enough,*

$$\mathbb{P}\Big(\max_{1\le i\le n} \nu\big(\hat{\mathbf{C}}_n(X_i) \triangle \mathbf{C}_P(X_i)\big) > A_\lambda c_{\gamma_n} \Big) = \mathcal{O}(n^{-\lambda}), \tag{4.9}$$

where c_{γ_n} *is such that* $h_{\gamma_n}/c_{\gamma_n}^2 \to 0$ *and* $h_{\gamma_n} \to 0$ *such that* $\gamma_n h_{\gamma_n}^{\ell+1}/\log(\gamma_n) \to \infty$.

Remark 1. It can be easily seen that, up to logarithmic factors, the rate of c_{γ_n} is $n^{-1/((\ell+3)(d+2))}$.

5. Non-compact manifolds

In this section, we discuss how to remove the compactness assumption imposed in Theorem 2, while retaining all the other hypotheses. We prove that the set $\hat{\mathbf{C}}_n(X_{n+1})$ satisfies *asymptotic conditional validity* (see Izbicki et al. (2022)). In other words, there exist random sets Λ_n such that

$$\mathbb{P}\Big(X_{n+1} \in \Lambda_n | \Lambda_n\Big) = 1 - o_{\mathbb{P}}(1),$$

and

$$\sup_{x_{n+1}\in\Lambda_n} \Big| \mathbb{P}\Big(Y_{n+1} \in \hat{\mathbf{C}}_n(X_{n+1}) \,\Big|\, X_{n+1} = x_{n+1}\Big) - \big(1-\alpha\big) \Big| = o(1).$$

According to Theorem 6 in Izbicki et al. (2022), to obtain asymptotic conditional validity it suffices to prove that

$$\mathbb{P}\Big(Y_{n+1} \in \hat{\mathbf{C}}_n(X_{n+1}) \triangle \mathbf{C}_P(X_{n+1})\Big) = o(1). \tag{5.10}$$

Theorem 3. *Assume H0 to H5. Let* $\hat{\mathbf{C}}_n(x)$ *be given by* (4.7) *and* $\mathbf{C}_P(x)$ *by* (2.1). *Then,* $\hat{\mathbf{C}}_n(X_{n+1})$ *is asymptotically conditionally valid.*

It is worth mentioning that asymptotic conditional validity is considerably weaker than (4.9) when $p(y|x)$ is bounded from above.

6. Computational aspects

A different method for estimating the conformal region is introduced in Izbicki et al. (2022). Instead of using a partition of $\mathbb{R}^d$ on cubes to get a partition of $X_1, \ldots, X_n$, to build kernel-based estimators, this approach employs a new partition constructed as follows. Consider, as in Izbicki et al. (2022), the functions

$$H(z|x) = \int_{\{y: p(y|x)\le z\}} p(y|x)\, d\nu(y), \qquad \hat{H}(z|x) = \int_{\{y:\hat{p}(y|x)\le z\}} \hat{p}(y|x)\, d\nu(y).$$

where $\hat{p}(y|x)$ is any estimator of $p(y|x)$, not necessarily kernel-based. Define the conditional α-quantile of $Y|X$ by $q_\alpha(x) = H^{-1}(\alpha|x)$. Similarly, let $\hat{q}_\alpha(x) = \hat{H}^{-1}(\alpha|x)$, be the estimate of the conditional α-quantile of $Y|X$.

Let $\mathcal{F}$ be a partition of $\mathbb{R}^+$. Then we create a partition $\mathcal{A} = \{A_1, \ldots, A_T\}$ of $X_1, \ldots, X_n$ by assigning X_i and X_j to the same set $A_r \in \mathcal{A}$ if and only if $\hat{q}_\alpha(X_i)$ and $\hat{q}_\alpha(X_j)$ lie in the same element of $\mathcal{F}$. We then build the kernel-based estimator (4.4) according to this new partition. This approach is particularly suitable for high-dimensional feature spaces.

We propose either the set $\hat{\mathbf{C}}_n(x)$ (as defined in (4.7)) or one of the algorithms presented in Izbicki et al. (2022). According to Theorem 25 in Izbicki et al. (2022), to establish the asymptotic conditional validity of this proposal (in the sense of (5.10)), it is necessary to assume some smoothness of $H(z, x)$ (see Assumption 23 in Izbicki et al. (2022)), as well as the compactness of M and the existence of sequences $\eta_n = o(1)$ and $\rho_n = o(1)$ satisfying

$$\mathbb{P}\left(\mathbb{E}\left[\sup_{y \in \mathcal{M}} \left(\hat{p}(y|X) - p(y|X)\right)^2 \,\middle|\, \hat{p} \right] \geq \eta_n \right) \leq \rho_n.$$

This condition indeed holds in our case; however, proving it in detail would require rewriting the proof of Theorem 1 in Cholaquidis et al. (2022). We leave this verification to the reader. As in the non-compact case, the consistency result obtained is considerably weaker than (4.9).

Since the computational burden of approximating $\hat{\mathbf{C}}_n(x)$ is high, Lei and Wasserman (2014) proposes replacing that set with $\mathbf{C}_n^+(x)$, see (6.11), which, by including $\hat{\mathbf{C}}_n(x)$, has coverage of at least $1 - \alpha$. Let $A_x \in \mathcal{A}$ be the element that contains x.

1. Let $\hat{p}(x, y)$ be the joint density estimator based on the subsample (of cardinality n_x) of points for which $X_i \in A_x$
2. Let $Z_i = (X_i, Y_i)$ for $i = 1, \ldots, n_x$, and let $Z_{(1)}, Z_{(2)}, \ldots, Z_{(n_x)}$ denote the sample ordered increasingly by the values $\hat{p}(X_i, Y_i)$.
3. Let $j = \lfloor n_x \alpha \rfloor$ and define

$$\mathbf{C}_n^+(x) = \left\{ y : \hat{p}(x, y) \geq \hat{p}\big(X_{(j)}, Y_{(j)}\big) - \frac{K(0)^2}{n_x\, h^{\ell+1}} \right\}. \tag{6.11}$$

The direct computation of the conformal set $\widehat{C}_n(x)$ based on the kernel estimator has computational cost $O(n^2)$, as each candidate y requires kernel evaluation across all samples. The simplified version $C_n^+(x)$ lowers this to $O(n \log n)$ when density estimates are precomputed on a grid or partition. In practice, datasets with several thousand observations can be handled efficiently on standard hardware. For high-dimensional manifolds (large d or ℓ), quantile-partition methods, Izbicki et al. (2022), or randomized nearest-neighbor schemes offer scalable subquadratic alternatives.

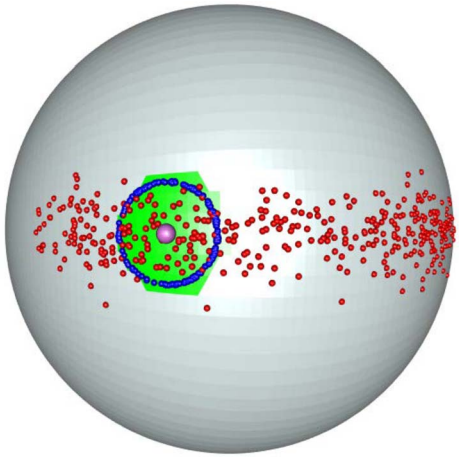

FIG 1. *Red: a sample* $Y_1, \ldots, Y_{400}$ *on* S^2. *Green: the estimated* 90%*-confidence set for* Y *if* $x = 0$. *Purple: the point prediction for* $x = 0$, *which is* $y = (1, 0, 0)$. *Blue dots belong to the boundary of the theoretical confidence set* (2.1) *for* $\alpha = 0.1$.

7. Simulation examples

7.1. Toy model: Output on the sphere

We consider a regression model with output variable denoted by Y_i defined on the unit sphere S^2. The input variable X_i takes values in the interval $[-1, 1]$. The model is given by the following probability distribution:

$$Y_i \sim \mathbb{F}\left(\frac{\eta + \beta X_i}{\|\eta + \beta X_i\|}, \kappa\right), \; i = 1, 2 \ldots, 400.$$

Here, $\mathbb{F}$ denotes the Von Mises–Fisher distribution (see Mardia and Jupp (2000)), and $X_i \sim \mathrm{U}(-1, 1)$ is an i.i.d. sample. For the simulation study, we set $\eta = (1, 0, 0)$, $\beta = (0, 0, 1)$, and $\kappa = 200$.

To estimate the kernel density, we use the estimator given by Equation (4.4) with a bandwidth parameter $h = 0.5$. Additionally, we partition the data into intervals $A_k = (-1 + (k-1)/2, -1 + k/2)$, where $k = 1, \ldots, 4$.

Using the proposed method, we obtain a 90% confidence set $\hat{\mathbf{C}}_{400}(0)$, shown in green in Figure 1. In this particular case, the unobserved output was $Y = (1, 0, 0)$, which is illustrated in purple in Figure 1. Points belonging to the boundary of the theoretical confidence set (2.1) for $\alpha = 0.1$ are displayed in blue.

7.2. Output in a Stiefel manifold

It is common to summarize information from a set of variables using Principal Component Analysis (PCA) or Factor Analysis. This approach is frequently

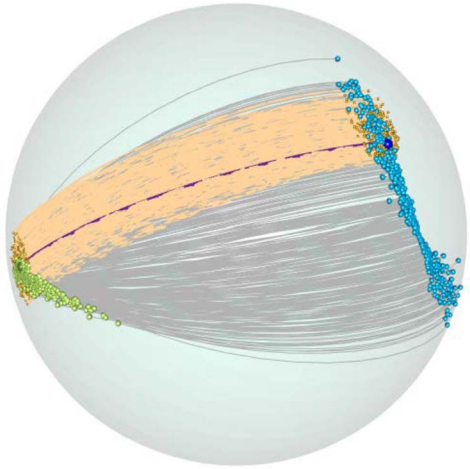

FIG 2. *A uniform sample of 10000 points in the* 95% *confidence set (orange arcs) for* $Y \in St(3,2)$ *if* $x = 0.1$. *The sample value for* $x = 0.1$ *is the* 3×2 *matrix,* y_x, *whose first column is* $(0.62, 0.5, 0.6)$ *and second column is* $(-0.3, 0.86, -0.44)$, *which is depicted in purple. Observed outputs are represented by gray arcs.*

employed in constructing indices within the social sciences, as demonstrated by Vyas (2006). Subsequently, researchers often attempt to explain these indices using other covariates. To illustrate this scenario, we provide a 'toy example' using simulations.

We will define a regression model whose output lies on the Stiefel Riemannian manifold $St(3,2)$, represented by 3×2 column-orthonormal matrices. The construction of the regression is as follows.

The input variable, denoted by X, is uniformly distributed on the interval $[0,1]$. For each value of X, we sample 400 points from a 3-dimensional Gaussian distribution centered at the origin $(0,0,0)$. The covariance matrix of this Gaussian distribution has eigenvalues 2, 1, and $1/2$, along with the corresponding eigenvectors $(2+X, 2+2X, 2-X)$, $(-2+X, 2-3X, -2+X)$, and $(1,0,0)$, respectively.

After generating these 400 points for each value of X, the output $Y \in St(3,2)$ is obtained by applying a PCA to these points and retaining only the first two principal directions. For the simulations analysis, we sample 500 points, denoted by $X_1, \ldots, X_{500}$.

In Figure 2, we display a sample of 500 output values, represented as gray arcs. The specific output value $y_{0.1} \in St(3,2)$, corresponding to the input $X = 0.1$, is depicted as a purple dotted arc.

To estimate the density, we use the estimator given by Equation (4.4), with a bandwidth parameter $h = 1$. Additionally, we divide the data into intervals $A_k = ((k-1)/5, k/5)$, where $k = 1, \ldots, 5$. It is important to note that the $St(3,2)$ data are embedded in $\mathbb{R}^6$ $(D = 6)$, and we consider the Euclidean distance in this space. The dimension of the submanifold in this case is $\ell = 3$.

To visualize the confidence set, we draw 10000 points within $St(3,2)$. The points falling within the confidence set are highlighted in orange in Figure 2.

8. Real-data examples

8.1. *Example 1: Regression between cylindrical manifolds with application to wind modeling*

In this study we address a problem of substantial practical relevance for the development of wind energy infrastructure in Uruguay. Specifically, the goal is to construct prediction sets for wind intensity R and direction θ, at a meteorological station located in Montevideo, based on simultaneous observations from a nearby station in Maldonado. These variables together form a point on a cylinder $\mathcal{D} = S^1 \times \mathbb{R}^+ \subset \mathbb{R}^3$, where S^1 represents the directional component (wind angle) and $\mathbb{R}^+$ the non-negative intensity. Thus, the regression problem consists of learning a mapping from one cylindrical manifold to another.

This task is not only of theoretical interest due to the non-Euclidean structure of the input and output spaces, but also of *strategic importance for national energy planning*. Accurate prediction of wind behavior at candidate sites is a critical step in the optimal placement of wind turbines, which are a major source of renewable energy in Uruguay's energy matrix.

We focus on moderate and extreme wind regimes (within the range 0 to 20 m/s), using hourly meteorological data recorded during the month of July between 2008 and 2016. The data were collected at two locations: the Laguna del Sauce Weather Station in Maldonado (Station 1) and the Carrasco International Airport Weather Station in Montevideo (Station 2), which are approximately 84 kilometers apart. To ensure reliability, only time points for which both stations reported valid measurements were retained, resulting in a total of 5,362 matched observations.

Our objective is to construct conformal prediction sets, for wind conditions at Station 2 given observations from Station 1. These sets aim to quantify predictive uncertainty while respecting the cylindrical geometry of the variables involved, offering robust tools for risk assessment and decision-making in wind farm siting.

Figures 3 and 4 show that there is a correlation in both wind direction and intensity between the two stations. The ξ-correlation, see Chatterjee (2021), between the intensities is 0.45. In the case of the directions, the angular correlation is 0.63. The angular correlation is the linear correlation between the variables $\sin\left(\theta_1 - \bar{\theta}_1\right)$ and $\sin\left(\theta_2 - \bar{\theta}_2\right)$, see Jammalamadaka and Sarma (1988).

In this example, the partition is

$$A_{i,j} = \left\{(\theta, R) \in \mathcal{D} : \theta \in [i\pi/4, (i+1)\pi/4] \text{ and } R \in [4j, 4(j+1)]\right\}$$

with $i = 0, 1, \ldots, 7$ and $j = 0, 1, \ldots, 4$. For the kernel density estimator we choose $h_k = 0.4$.

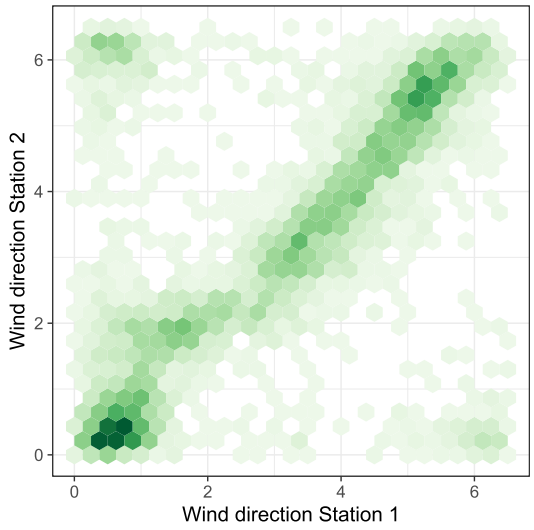

FIG 3. *Scatter plot between the angular directions (in radians) of the winds recorded at Station 1 and Station 2 in the months of July between 2008 and 2016.*

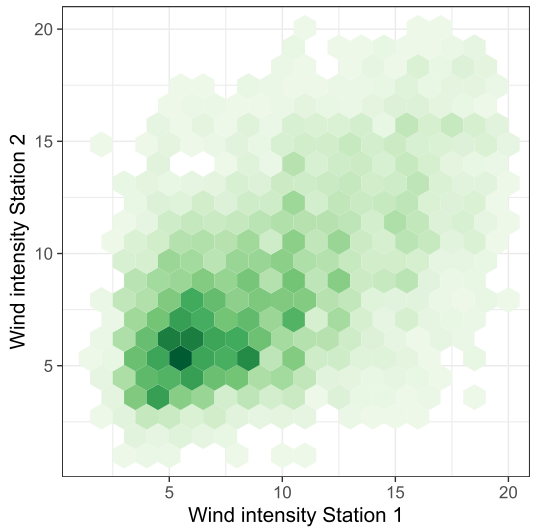

FIG 4. *Scatter plot between the intensities (in m/s) of the winds recorded at Station 1 and Station 2 in the months of July between 2008 and 2016.*

Figure 5 shows the confidence set (at 80%) obtained by our method for $(\theta_1, R_1) = (2.3 \text{ radians}, 5.1\text{m/s})$. This was recorded at Station 1 on 2008-07-01 at 7 pm. On the same date and time the data recorded at Station 2 were $(\theta_2, R_2) = (2.4 \text{ radians}, 6.6\text{m/s})$ (depicted in purple in Figure 5).

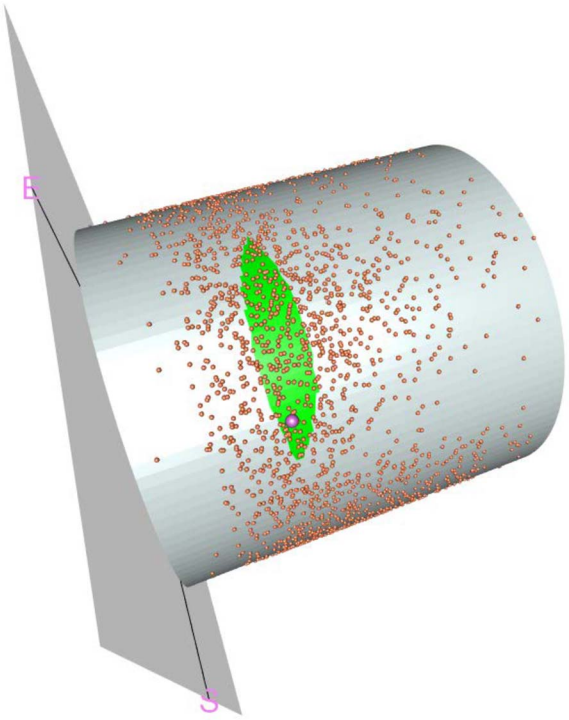

FIG 5. *Winds (not exceeding 20 m/s) at Station 2 in the months of July between 2008 and 2016 (red). The green area is the 80% confidence set for the data recorded at Station 1 of* $(2.3\ \textit{radians}, 5.1\ \textit{m/s})$ *on 2008-07-01 at 7 pm. The purple dot represents the wind recorded at the same time at Station 2* $((\theta_2, R_2) = (2.4\ \textit{radians}, 6.6\ \textit{m/s}))$.

8.2. *Example 2: Regression on the simplex with application to multiclass classification*

Regression models whose outputs lie in the $(d-1)$-dimensional unit simplex Δ^{d-1} play an essential role in many applied contexts (see, for example, (Aitchison, 1986; Pawlowsky-Glahn et al., 2015)). In supervised classification, a variety of machine-learning algorithms produce vectors of class-membership probabilities—i.e., elements of Δ^{d-1}—which we denote by Y_i. Here, each Y_i is obtained by using Random Forest (RF). Because these vectors reside in Δ^{d-1}, their components naturally estimate the probability of membership in each class. The final class prediction is then chosen as the one with the highest estimated probability, equivalent to identifying the Voronoi cell of the simplex in which Y_i falls (see Figure 6).

Our goal is to construct a $100(1-\alpha)\%$ confidence region for observations on the simplex Δ^{d-1} by adapting Algorithm 1 from Izbicki et al. (2022) (CD-split) to this setting. In particular, we follow Izbicki and Lee (2017) to estimate the conditional density, employing an appropriate Fourier-type basis on the

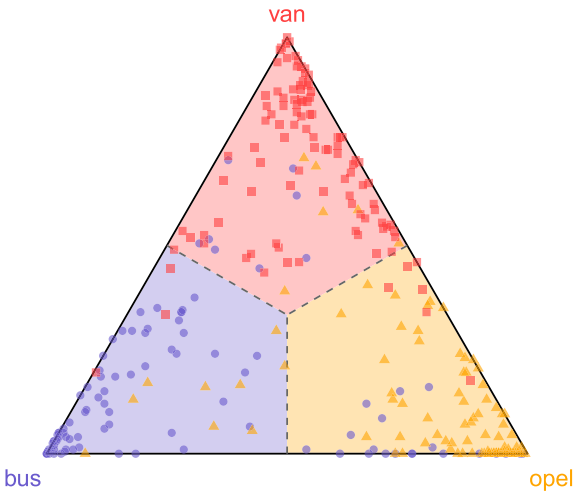

FIG 6. *Test set predictions for vehicle classification. Colors denote the true class labels, and Voronoi regions indicate the predicted class according to Random Forest.*

simplex—such as the Bernstein polynomial basis (Farouki et al., 2003; Ghosal, 2001).

The intersection of this confidence region with the Voronoi tessellation of Δ^{d-1} yields the *conformal class set*, i.e., the subset of classes that are statistically consistent with the observation at level α. Moreover, this approach provides insight into how the predictive uncertainty is distributed among the classes, by analyzing the proportion of the confidence region that falls within each Voronoi cell.

An application of the proposed methodology is carried out on a real-world dataset, where the covariate space is 5-dimensional and the response variable lies on Δ^2, representing class probabilities, obtained, as stated, by means of RF. To this end, we consider the publicly available `Vehicle` dataset from the `mlbench` package in `R` (Leisch et al., 2007).

The `Vehicle` dataset originates from a benchmark study conducted at the Turing Institute in the early 1990s. It was designed to evaluate the performance of classification algorithms in distinguishing between different types of road vehicles based on geometric features extracted from digitized images of their silhouettes. Each observation in the dataset corresponds to a vehicle image and is characterized by 5 numerical covariates, such as aspect ratio, edge count, and moment-based shape descriptors.

The categorical response variable indicates the type of vehicle. For the purposes of this study, we restrict attention to a subset of the data containing three of the four classes (`bus`, `opel`, and `van`), so that the associated output probabilities lie on Δ^2. This structure is ideal for illustrating our simplex-based predictive modeling approach. The dataset contains a total of 628 observations, and its moderate size, combined with its multi-class nature, makes it well-suited

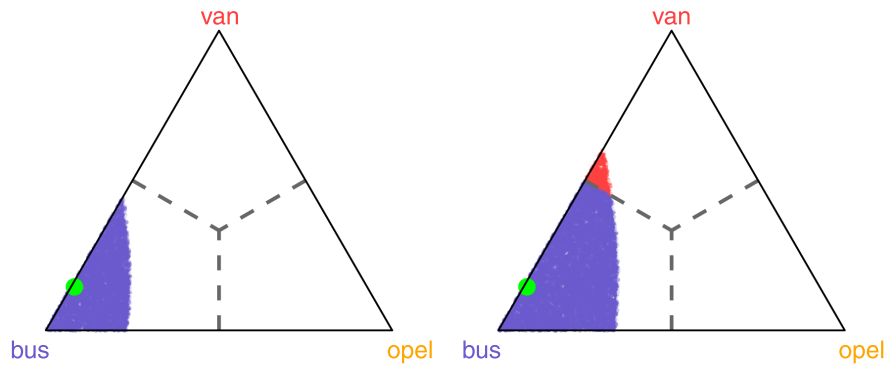

FIG 7. *Left panel:* 90% *conformal prediction region; right panel:* 95% *conformal prediction region for* $y_{n+1} = (0.845,\ 0.01,\ 0.145)$ *(shown in green). The background colors represent the Voronoi regions of the predicted classes over the 2-dimensional simplex* Δ^2*. Classification was performed using a Random Forest model.*

for benchmarking methods involving compositional or probabilistic responses.

The dataset is randomly split into two equal parts: 50% of the observations are used for training the model, and the remaining 50% are reserved as a test set for evaluating predictive performance.

Figure 7 displays the test data projected onto Δ^2. Each point corresponds to a predicted probability vector, and its color indicates the true class label. The predicted class for a point is the Voronoi cell that contains the point.

We assess confidence levels of 90% and 95% for the covariate vector $\mathbf{x}_{n+1} = (84, 45, 66, 150, 65)$. The corresponding class probabilities predicted by the Random Forest are $\mathbf{y}_{n+1} = (0.845, 0.01, 0.145)$, highlighted in purple in Figure 7. The resulting class band, restricted to the "bus" and "van" categories, offers a detailed decomposition of predictive uncertainty by indicating the fraction of the confidence region attributable to each class. At the 95% confidence level, approximately 93% of the band resides within the "bus" region, while the remaining 7% falls within the "van" region.

9. Conclusions and future work

We work within the conformal-inference framework, where the response variable takes values on a Riemannian manifold while the covariates lie in $\mathbb{R}^d$. Under assumptions analogous to those in Lei and Wasserman (2014), we extend their results to show that the conditional oracle set $\mathbf{C}_P(x)$ can be estimated consistently. A central tool in our analysis is the almost-sure, uniform consistency of the classical kernel density estimator on manifolds, as proved in Cholaquidis et al. (2022). We also cover the cases of manifolds with boundary and non-compact manifolds, and we discuss computational strategies for handling high-dimensional input data.

Our theoretical contributions are demonstrated through two simulation studies and two real-world data examples.

As a promising direction for future work, we plan to further investigate the approach proposed by Izbicki et al. (2022) for prediction tasks involving dependent variables defined on Riemannian manifolds. In particular, we aim to build upon the ideas discussed by Dheur et al. (2025), who present a comprehensive comparative study of multivariate conformal prediction methods and introduce novel conformity scores that extend univariate approaches to multi-output regression. Their framework provides valuable insights into the challenges of modeling output dependencies and computational efficiency in multivariate settings. Integrating these ideas could pave the way toward more advanced and geometrically informed conformal prediction models for complex data domains.

Finally, one could explore alternative nonconformity scores beyond the KDE-based criterion we employ—for example, quantile-based or regression-based scores—each of which demands a nontrivial adaptation of the Euclidean theory to the manifold setting.

Appendix A

Proof of Theorem 1. Since $n_k h_{n_k}^{\ell+3}/\log(n_k) \to \infty$, then, from theorem 1 of Cholaquidis et al. (2022), for all $\epsilon > 0$, with probability one, for all n large enough,

$$\frac{c_{n_k}}{h_{n_k}} \sup_{y \in \mathcal{M}_{n_k}} |\hat{p}(y|A_k) - p(y|A_k)| \leq \epsilon. \tag{A.12}$$

Note that the index n_0 for which (A.12) holds depends on the choice of k. However, from (4.5), we know that with probability one, $n_k \geq \gamma_n$ for all k when n is sufficiently large. Since $h_n/c_n \to 0$ monotonically, it follows that for all k, $c_{n_k}/h_{n_k} \leq c_{\gamma_n}/h_{\gamma_n}$. Hence, we can conclude that (4.8) holds for all k. □

A.1. Proof of Theorem 2

The proof of Theorem 2 is based on some technical lemmas.

Lemma 1. *Under the hypotheses of Theorem 1, assume also H0 to H3. Let* $R_n(x) = \sup_{y \in \mathcal{M}_{\gamma_n}} |\hat{p}(y|A_k) - p(y|x)|$. *Then, with probability one, for n sufficiently large, we have*

$$\sup_{x \in supp(P_X)} R_n(x) \leq h_{\gamma_n}/c_{\gamma_n} + L w_n \sqrt{d}, \tag{A.13}$$

where $\gamma_n = \lfloor b_1 n w_n^d/2 \rfloor$ *and* $\inf_{x \in M_{\gamma_n}} \rho(x, \partial\mathcal{M}) = c_{\gamma_n}$.

Proof. From (4.5), we can assume that n is large enough and such that for all k, $b_1 n w_n^d \leq n_k \leq b_2 n w_n^d$. From (4.8), we obtain with probability one for n large enough,

$$\sup_{y \in \mathcal{M}_{\gamma_n}} |\hat{p}(y|A_k) - p(y|A_k)| \leq h_{n_k}/c_{n_k}. \tag{A.14}$$

Using assumption H3 and (4.5), we have

$$\begin{aligned}\sup_{x\in\mathrm{supp}(P_X)} R_n(x) &\le \sup_{x\in\mathrm{supp}(P_X)} \sup_{y\in\mathcal{M}_{\gamma_n}} |\hat{p}(y|A_k) - p(y|x)| \\ &\le \sup_{x\in\mathrm{supp}(P_X)} \sup_{y\in\mathcal{M}_{\gamma_n}} |\hat{p}(y|A_k) - p(y|A_k)| \\ &\quad + \sup_{y\in\mathcal{M}_{\gamma_n}} |p(y|A_k) - p(y|x)| \\ &\le h_{\gamma_n}/c_{\gamma_n} + Lw_n\sqrt{d}.\end{aligned}$$

Thus, we have shown (A.13). □

Lemma 2. *Assume H0 to H4. Let $w_n = (\log(n)/n)^{1/(d+2)}$, and assume that γ_n and h_n are as in Lemma 1. Then, there exists $C > 0$ such that with probability one, for n large enough*

$$\sup_k |\hat{p}(\cdot|A_k)|_\infty \le C.$$

Proof. Recall that we are assuming K to be a Gaussian kernel. We assume that n is large enough so that (4.5) holds. Following the proof of theorem 1 of Cholaquidis et al. (2022), we define

$$W_j(y) = (1/(n_k h_{n_k}^\ell))\mathbf{1}_{\{X_j\in A_k\}} K(\|Y_j - y\|/h_{n_k}),$$

$V_j(y) = W_j(y) - \mathbb{E}(W_j(y))$, and $S_n(y) = \sum_{j=1}^n V_j(y)$. Let $\eta > \ell+1$ and consider a covering $B(p_1, h_{n_k}^\eta), \ldots, B(p_l, h_{n_k}^\eta)$ of $\mathcal{M}$. Then,

$$\sup_y |S_n(y)| \le \max_{1\le j\le l} \sup_{y\in B(p_j,h_{n_k}^\eta)} \frac{1}{n_k h_{n_k}^\ell}|S_n(y) - S_n(p_j)| + \max_{1\le j\le l} \frac{1}{n_k h_{n_k}^\ell}|S_n(p_j)| = I_1 + I_2. \quad \text{(A.15)}$$

Since K is Lipschitz, we have $I_1 \le c h_{n_k}^{\eta-\ell-1}$ for some constant $c > 0$. Applying Bernstein's inequality to $S_n(p_j)$ for fixed p_j and A_k, we get

$$\mathbb{P}\Big(\frac{1}{n_k h_{n_k}^\ell}|S_n(p_j)| > \epsilon\Big) \le c_1 \exp(-c_2 n_k h_{n_k}^\ell \epsilon^2),$$

where c_1 and c_2 are positive constants.

Then, if $\kappa_n = \lfloor 3b_2 n w_n^d \rfloor$,

$$\mathbb{P}\Big(\sup_k \sup_y |S_n(y)| > \epsilon\Big) \le c_3 w_n^{-d} \exp(-c_3 n w_n^d h_{\kappa_n}^\ell).$$

By the Borel–Cantelli lemma, it follows that $\sup_k \sup_y |S_n(y)| \to 0$ almost-surely.

Conditioning on $X_1, \ldots, X_n$, we have

$$\mathbb{E}\big[W_j(y) \,\big|\, X_1, \ldots, X_n\big] = \frac{\mathbf{1}_{\{X_j\in A_k\}}}{n_k\, h_{n_k}^\ell} \int_{\mathcal{M}} K\Big(\frac{\|z-y\|}{h_{n_k}}\Big)\, p(z \,|\, X_j)\, d\nu(z),$$

By H3, set $M_p := \sup_{x\in\mathrm{supp}(P_X)} \|p(\cdot\,|\,x)\|_\infty < \infty$. Therefore,

$$\mathbb{E}\big[W_j(y)\,\big|\,X_1,\dots,X_n\big] \le \frac{\mathbf{1}_{\{X_j\in A_k\}}}{n_k}\, M_p \underbrace{\int_{\mathcal{M}} \frac{1}{h_{n_k}^{\ell}} K\!\left(\frac{\|z-y\|}{h_{n_k}}\right) d\nu(z)}_{=:I(y,h_{n_k})}.$$

We show that $\sup_{y\in\mathcal{M},\ h\in(0,1]} I(y,h) \le C_0$ for some constant C_0.

Fix $y\in\mathcal{M}$. Let $r\in(0,\mathrm{inj}(\mathcal{M})/2]$ and split $\mathcal{M} = B_\rho(y,r)\cup(\mathcal{M}\setminus B_\rho(y,r))$, where $B_\rho(y,r)$ is the geodesic ball. Because $\mathcal{M}$ is a compact $\mathcal{C}^2$ submanifold, there exist constants $c_1,c_2\in(0,\infty)$ and r small enough such that, for all ξ with $\|\xi\|\le r$, $c_1\|\xi\| \le \big\|\exp_y(\xi)-y\big\| \le c_2\|\xi\|$. Moreover, the Jacobian satisfies $J_y(\xi) = 1+O(\|\xi\|^2)$ uniformly in y. On $B_\rho(y,r)$,

$$\int_{B_\rho(y,r)} \frac{1}{h^\ell} K\!\left(\frac{\|z-y\|}{h}\right) d\nu(z) = \int_{\|\xi\|\le r} \frac{1}{h^\ell} K\!\left(\frac{\|\exp_y(\xi)-y\|}{h}\right) J_y(\xi)\, d\xi.$$

Therefore,

$$\int_{\|\xi\|\le r} \frac{1}{h^\ell} K\!\left(\frac{\|\exp_y(\xi)-y\|}{h}\right) J_y(\xi)\, d\xi \;\le\; \int_{\|\xi\|\le r} \frac{1}{h^\ell} K\!\left(\frac{c_1\|\xi\|}{h}\right) \big(1+C\|\xi\|^2\big)\, d\xi.$$

With the change of variables $u = (c_1/h)\,\xi$, we get

$$\begin{aligned}\int_{\|\xi\|\le r} \frac{1}{h^\ell} K\!\left(\frac{c_1\|\xi\|}{h}\right) \big(1+C\|\xi\|^2\big)\, d\xi &= \frac{1}{c_1^\ell}\int_{\|u\|\le r/h} K(\|u\|)\left(1+C\,\frac{h^2}{c_1^2}\|u\|^2\right) du \\ &\le\; C_1,\end{aligned}$$

where C_1 is independent of y and h. On $\mathcal{M}\setminus B_\rho(y,r)$ we have $\|z-y\|\ge r/c_2$, hence

$$\int_{\mathcal{M}\setminus B_\rho(y,r)} \frac{1}{h^\ell} K\!\left(\frac{\|z-y\|}{h}\right) d\nu(z) \le \frac{1}{h^\ell}\, K\!\left(\frac{r}{c_2 h}\right) \nu(\mathcal{M}). \tag{A.16}$$

For a Gaussian kernel K, $K(t)\le Ce^{-at^2}$, so $h^{-\ell}K(r/(c_2h)) \le C\,h^{-\ell}e^{-ar^2/(c_2^2h^2)}$. The function $u\mapsto u^\ell e^{-ar^2u^2/c_2^2}$ (with $u=1/h$) is bounded on $(0,\infty)$, hence there exists C_2 such that

$$\sup_{h\in(0,1]} \frac{1}{h^\ell}\, K\!\left(\frac{r}{c_2 h}\right) \;\le\; C_2.$$

Thus, (A.16) is bounded from above by $C_2\,\nu(\mathcal{M})$.

Combining (i) and (ii) gives $I(y,h)\le C_0 := C_1 + C_2\,\nu(\mathcal{M})$, uniformly in $y\in\mathcal{M}$ and $h\in(0,1]$. Therefore,

$$\mathbb{E}\big[W_j(y)\,\big|\,X_1,\dots,X_n\big] \le \frac{\mathbf{1}_{\{X_j\in A_k\}}}{n_k}\, M_p\, C_0 \;\le\; \frac{C}{n_k},$$

with $C := M_pC_0$. □

Let us define $L_x(t) := \{y : p(y|x) \geq t\}$, $L_x^l(t) := \{y : p(y|x) \leq t\}$. Lemma 6 of Lei and Wasserman (2014) proves a slightly modified version of the following lemma.

Lemma 3. *Assume H0 to H5. Let* $\gamma_n = \lfloor b_1 n w_n^d/2 \rfloor$, $h_{\gamma_n} \to 0$, *and* $\mathcal{M}_{\gamma_n} \subset \mathcal{M}$ *be a sequence of closed sets such that* $\inf_{x\in\mathcal{M}_{\gamma_n}} \rho(x, \partial\mathcal{M})/h_{\gamma_n} \to \infty$. *Then, for any* $\lambda > 0$, *there exists* $\xi_{2,\lambda}$ *such that, for* n *large enough,*

$$\mathbb{P}\left(\max_{1\leq i\leq n} V_n(X_i) > \xi_{2,\lambda} w_n \right) = \mathcal{O}(n^{-\lambda}), \tag{A.17}$$

where $V_n(x) = \sup_{t\geq t_0} |\hat{P}(L_x^l(t)|A_k) - P(L_x^l(t)|x)|$. *Here,* $\hat{P}(\cdot|A_k)$ *is the empirical distribution of* $Y|X \in A_k$, t_0 *is given in H4, and* $w_n = (\log(n)/n)^{1/(d+2)}$.

Proof. We will provide a sketch of the proof of this lemma since it follows essentially the same idea used to prove lemma 6 of Lei and Wasserman (2014).

Let $x \in A_k$ be fixed. Note that $\{L_x^l(t) : t \geq t_0\}$ is a nested class of sets with Vapnik–Chervonenkis dimension 1. A standard VC tail bound yields $\mathbb{P}\big(\sup_t |\hat{P}(L_x^l(t)\,|\,A_k) - P(L_x^l(t)\,|\,A_k)| > \epsilon\big) \leq C_1 \exp(-C_2 n_k \epsilon^2)$. Taking $\epsilon = Bw_n$ and using $n_k \gtrsim n w_n^d$ on $\mathcal{E}_n$ gives

$$\mathbb{P}\left(\sup_t \left| \hat{P}(L_x^l(t)|A_k) - P(L_x^l(t)|A_k) \right| > Bw_n \right) = \mathcal{O}(n^{-\lambda}),$$

where we used that $w_n = (\log(n)/n)^{1/(d+2)}$. On the other hand,

$$|P(L_x^l(t)|A_k) - P(L_x^l(t)|x)| \leq L w_n \nu(L_x(t))\sqrt{d} \leq L w_n \nu(L_x(t_0))\sqrt{d} \leq L\nu(\mathcal{M})\sqrt{d} w_n.$$

Let $x' \in A_k$. Then

$$|\hat{P}(L_{x'}^l|A_k) - P(L_{x'}^l(t)|x')| \leq |\hat{p}(\cdot|A_k)|_\infty \nu(L_x^l(t) \triangle L_{x'}^l(t)) + V_n(x) + |G_x(t) - G_{x'}(t)|$$

where $G_x(t) = P(L_x^l(t)|x)$. From (30) in Lei and Wasserman (2014),

$$|G_x(t) - G_{x'}(t)| \leq c_3 w_n^{1\wedge\gamma}.$$

Here, c_3 is a positive constant and γ is as in H4. From (29) in Lei and Wasserman (2014), $\nu(L_x^l(t) \triangle L_{x'}^l(t)) \leq c_4 w_n^\gamma$ for some positive constant c_4. Lastly, $\sup_k |\hat{p}(\cdot|A_k)|_\infty$ is bounded, for n large enough, using Lemma 2. □

We recall lemma 8 of Lei and Wasserman (2014).

Lemma 4. *Fix* $\alpha > 0$, $t_0 > 0$ *and* $\epsilon > 0$. *Suppose that* p *is a density function that satisfies H4, and* $\hat{p}$ *an estimator such that* $|\hat{p} - p|_\infty < v_1$. *Let* $\hat{P}$ *be a probability measure satisfying* $\sup_{t\geq t_0} |\hat{P}(L^l(t)) - P(L^l(t))| < v_2$. *Define*

$$\hat{t}^\alpha = \inf\{t \geq 0 : \hat{P}(\hat{L}^l(t)) \geq \alpha\}.$$

Assume that v_1 *and* v_2 *are sufficiently small so that* $v_1 + c_1^{1/\gamma} v_2^{1/\gamma} \leq t^\alpha - t_0$ *and* $c_1^{-1/\gamma} v_2^{1/\gamma} \leq \epsilon_0$ *where* c_1 *and* γ *are the constants given in H4. Then*

$$|\hat{t}^\alpha - t^\alpha| \leq v_1 + c_1^{-1/\gamma} v_2^{1/\gamma}$$

Moreover, for any $\tilde{t}^\alpha$ *such that* $|\tilde{t}^\alpha - \hat{t}^\alpha| \leq v_3$*, if* $2v_1 + c_1^{1/\gamma} v_2^{1/\gamma} + v_3 \leq \epsilon_0$*, then there are constants* ξ_1, ξ_2 *and* ξ_3 *such that* $\nu(\hat{L}(\tilde{t}^\alpha) \triangle L(t^\alpha)) \leq \xi_1 v_1 + \xi_2 v_2 + \xi_3 v_3$.

Proof of Theorem 2

In what follows we assume that n is sufficiently large to satisfy (4.5). We will consider a sequence of compact sets $\mathcal{M}_n \subset \mathcal{M}$ such that $\inf_{x \in \mathcal{M}_n} \rho(x, \partial\mathcal{M}) = c_n > 0$ and $\sup_{x \in \partial\mathcal{M}} \rho(x, \mathcal{M}_n) \leq 2c_n$, where c_n is chosen such that $h_{\gamma_n}/c_{\gamma_n}^2 \to 0$.

Throughout the proof we denote by a a generic positive constant. Write

$$\nu(\hat{\mathbf{C}}_n(x) \triangle \mathbf{C}_P(x)) = \nu(\hat{\mathbf{C}}_n(x) \triangle \mathbf{C}_P(x) \cap \mathcal{M}_{\gamma_n}) + \nu\Big(\hat{\mathbf{C}}_n(x) \triangle \mathbf{C}_P(x) \cap (\mathcal{M} \setminus \mathcal{M}_{\gamma_n})\Big)$$

where γ_n is as in Lemma 3 and $\mathcal{M}_{\gamma_n} \subset \mathcal{M}$ is a sequence of closed sets such that $\inf_{x \in \mathcal{M}_{\gamma_n}} \rho(x, \partial\mathcal{M}) = c_{\gamma_n}$ satisfies $h_{\gamma_n}/c_{\gamma_n} \to 0$.

We will first prove that there exists a $a > 0$ such that

$$\nu(\mathcal{M} \setminus \mathcal{M}_{\gamma_n}) \leq a c_{\gamma_n}. \tag{A.18}$$

Since $\mathcal{M}$ is $\mathcal{C}^2$, it has positive reach, denoted by $\tau_{\mathcal{M}}$, as shown in proposition 14 of Thäle (2008). Let $A_r = \{x \in \mathcal{M} : \rho(x, \partial\mathcal{M}) < r\}$. From

$$\sup_{x \in \partial\mathcal{M}} \rho(x, \mathcal{M}_n) \leq 2c_n,$$

it follows that $\mathcal{M} \setminus \mathcal{M}_{\gamma_n} \subset A_{2c_{\gamma_n}}$.

We write $m_n = 2\tau_{\mathcal{M}} \sin(c_{\gamma_n}/\tau_{\mathcal{M}})$, and from proposition A.1 in Aamari et al. (2023), we know that $\Upsilon_{\gamma_n} := \{B(x, m_n) : x \in \partial\mathcal{M}\}$ covers $A_{2\gamma_n}$. Since $\partial\mathcal{M}$ has finite Minkowski content due to its positive reach (see corollary 3 of Ambrosio et al. (2008)), and $\sin(x) \approx x$ when $x \to 0$, there exists a $a > 0$ such that $|\Upsilon_{\gamma_n}|/c_{\gamma_n}^{D-\ell-1} \leq a$ for all sufficiently large n, where $|\Upsilon_{\gamma_n}|$ is the D-dimensional Lebesgue measure of Υ_{γ_n}.

Thus, $A_{2c_{\gamma_n}}$ can be covered by at most $a c_{\gamma_n}^{-\ell+1}$ balls of radius $2m_n$ centered at $\partial\mathcal{M}$, and the ν-measure of each of these balls is bounded from above by $B c_{\gamma_n}^{\ell}$ by corollary 1 of Aaron and Cholaquidis (2020). From this, A.18 follows.

Now, to bound $\nu(\hat{\mathbf{C}}_n(x) \triangle \mathbf{C}_P(x) \cap \mathcal{M}_{\gamma_n})$, we follow the approach used in the proof of theorem 1 in Lei and Wasserman (2014). We apply Lemma 4 to the density function $p(y|x)$ and the empirical measure $\hat{P}(\cdot|A_k)$, as well as the estimated density function $\hat{p}(\cdot|A_k)$. Here, we provide a sketch of the main changes made to the proof. We denote by $\hat{L}$ the upper level set of $\hat{p}(\cdot|A_k)$.

Let $\{i_1, \ldots, i_{n_k}\} = \{i : 1 \leq i \leq n, X_i \in A_k\}$. From lemma 3 in Lei and Wasserman (2014), conditioning on $(i_1, \ldots, i_{n_k})$,

$$\hat{L}\Big\{\hat{p}(X_{(i_\alpha)}, Y_{(i_\alpha)}|A_k)\Big\} \subset \hat{\mathbf{C}}_n(x) \subset \hat{L}\Big\{\hat{p}(X_{(i_\alpha)}, Y_{(i_\alpha)}|A_k) - (n_k h_{n_k})^{-1} \psi_K\Big\},$$

where $\psi_K = \sup_{x,x'} |K(x) - K(x')|$ with $(X_{(i_\alpha)}, Y_{(i_\alpha)})$ is the element of

$$\{(X_{1_1}, Y_{i_1}), \ldots, (X_{i_{n_k}}, Y_{i_{n_k}})\}$$

such that $\hat{p}(Y_{i_\alpha}|A_k)$ ranks $\lfloor n_k \alpha \rfloor$. Let $\hat{t}^\alpha = \hat{p}(X_{(i_\alpha)}, Y_{(i_\alpha)})$. It is easy to check that

$$\hat{t}^\alpha = \inf\{t \geq 0 : \hat{P}(L^l(t)|A_k) \geq \alpha\}.$$

Consider the event

$$E = \Big\{ \sup_x R_n(x) \leq 2\frac{h_{\gamma_n}}{c_{\gamma_n}},\ \sup_x V_n(x) \leq \xi_{2,\lambda} w_n \Big\}.$$

From $n w_n^d h_{\gamma_n}^{\ell+3} / \log(\gamma_n) \to \infty$, it follows that $w_n / h_{\gamma_n} \to 0$. Then from Lemmas 1 and 3, $\mathbb{P}(E^c) = \mathcal{O}(n^{-\lambda})$. Let $r_n := h_{\gamma_n}/c_{\gamma_n}$. Since $w_n/h_{\gamma_n} \to 0$, then the event

$$E_1 = \Big\{ \sup_x R_n(x) \leq 2r_n, \sup_x V_n(x) \leq \xi_{2,\lambda} r_n \Big\}$$

is such that for all n large enough, $\mathbb{P}(E_1^c) = \mathcal{O}(n^{-\lambda})$. From Lemma (4) with $v_1 = 2r_n$ and $v_2 = \xi_{2,\lambda} r_n$, we obtain that, for n large enough,

$$\mathbb{P}\Big(\sup_x \nu\Big(\hat{L}(\hat{t}^\alpha) \triangle L_x(t^\alpha) \cap \mathcal{M}_{\gamma_n} \Big) \geq \xi_\lambda r_n \Big) = \mathcal{O}(n^{-\lambda}). \tag{A.19}$$

Let $\tilde{t}^\alpha = \hat{t}^\alpha - (\gamma_n h_{\gamma_n})^{-1} \psi_K$ and $v_3 = (\gamma_n h_{\gamma_n})^{-1} \psi_K$. From $n w_n^d h_{\gamma_n}^{\ell+3} / \log(\gamma_n) \to \infty$, it follows that $v_3 \to 0$. Applying Lemma 4, we get that

$$\mathbb{P}\Big(\sup_x \nu\Big(\hat{L}(\tilde{t}^\alpha) \triangle L_x(t^\alpha) \cap \mathcal{M}_{\gamma_n} \Big) \geq \xi_\lambda^1 r_n + \xi_\lambda^2 v_3 \Big) = \mathcal{O}(n^{-\lambda}).$$

Since $v_3 \leq r_n$ for n large enough, we get that

$$\mathbb{P}\Big(\sup_x \nu\Big(\hat{L}(\tilde{t}^\alpha) \triangle L_x(t^\alpha) \cap \mathcal{M}_{\gamma_n} \Big) \geq 2\xi_\lambda^1 r_n \Big) = \mathcal{O}(n^{-\lambda}). \tag{A.20}$$

Lastly, we write

$$\nu(\hat{\mathbf{C}}_n(x) \triangle \mathbf{C}_P(x) \cap \mathcal{M}_{\gamma_n}) \leq \nu\Big(\hat{L}(\hat{t}^\alpha) \triangle L_x(t^\alpha) \cap \mathcal{M}_{\gamma_n} \Big) + \nu\Big(\hat{L}(\tilde{t}^\alpha) \triangle L_x(t^\alpha) \cap \mathcal{M}_{\gamma_n} \Big)$$

and then the theorem follows from (A.19) and (A.20).

Proof of Theorem 3

To prove (5.10), fix any $\epsilon > 0$ and choose a sufficiently large compact smooth submanifold $\mathcal{K} = \mathcal{K}(\epsilon) \subset \mathcal{M}$ fulfilling H0, such that $\mathbb{P}(Y \in \mathcal{K}^c) < \epsilon$. The existence of such a submanifold follows from the existence of smooth exhaustion functions (see Proposition 2.28 in Lee (2013)). Then, by a simple union bound,

$$\mathbb{P}\Big(Y_{n+1} \in \hat{\mathbf{C}}_n(X_{n+1}) \triangle \mathbf{C}_P(X_{n+1})\Big) \leq \\ \mathbb{P}\Big(Y_{n+1} \in \Big(\hat{\mathbf{C}}_n(X_{n+1}) \triangle \mathbf{C}_P(X_{n+1})\Big) \cap \mathcal{K}\Big) + \epsilon.$$

Now, fixing $\mathcal{K}$, we take—as in the proof of Theorem 2—a sequence of closed sets $\mathcal{K}_{\gamma_n} \subset \mathcal{K}$ such that

$$\inf_{x \in \mathcal{K}_{\gamma_n}} \rho(x, \partial\mathcal{K}) = c_{\gamma_n} \to 0 \quad \text{and} \quad \frac{h_{\gamma_n}}{c_{\gamma_n}} \to 0 \quad \text{as } n \to \infty.$$

Then we decompose $\mathbb{P}\Big(Y_{n+1} \in \hat{\mathbf{C}}_n(X_{n+1}) \triangle \mathbf{C}_P(X_{n+1}) \cap \mathcal{K}\Big)$ equals

$$\mathbb{P}\Big(Y_{n+1} \in \big(\hat{\mathbf{C}}_n(X_{n+1}) \triangle \mathbf{C}_P(X_{n+1})\big) \cap \mathcal{K}_{\gamma_n}\Big) + \\ \mathbb{P}\Big(Y_{n+1} \in \big(\hat{\mathbf{C}}_n(X_{n+1}) \triangle \mathbf{C}_P(X_{n+1})\big) \cap \big(\mathcal{K} \setminus \mathcal{K}_{\gamma_n}\big)\Big).$$

Since $\mathcal{K}$ is compact and, for all x, $p(y|x)$ is a continuous function of y, for n large enough we have

$$\mathbb{P}\Big(Y_{n+1} \in \big(\hat{\mathbf{C}}_n(X_{n+1}) \triangle \mathbf{C}_P(X_{n+1})\big) \cap \big(\mathcal{K} \setminus \mathcal{K}_{\gamma_n}\big)\Big) < \epsilon.$$

Finally, the proof that

$$\mathbb{P}\Big(Y_{n+1} \in \big(\hat{\mathbf{C}}_n(X_{n+1}) \triangle \mathbf{C}_P(X_{n+1})\big) \cap \mathcal{K}_{\gamma_n}\Big) \to 0$$

follows by using the continuity of $p(y|x)$, that $\mathcal{K}$ is in the hypotheses of Theorem 2, and the fact established in the proof of Theorem 2 that

$$\nu\Big((\hat{\mathbf{C}}_n(x) \triangle \mathbf{C}_P(x)) \cap \mathcal{K}_{\gamma_n}\Big) \to 0$$

as $n \to \infty$.

Acknowledgment

Authors are grateful to Tyrus Berry, for his insightful comments on the results obtained in his work with Timothy Sauer. We warmly thank the two anonymous referees for their thoughtful comments and valuable suggestions, which have significantly improved the quality of the paper.

Funding

The research of the first and third authors has been partially supported by grant FCE-3-2022-1-172289 from ANII (Uruguay), 22MATH-07 form MATH – AmSud (France-Uruguay) and 22520220100031UD from CSIC (Uruguay). Support from the ANR-3IA Artificial and Natural Intelligence Toulouse Institute is gratefully acknowledged.

References

Aamari, E., C. Aaron, and C. Levrard (2023). Minimax boundary estimation and estimation with boundary. *Bernoulli 29*(4), 3334–3368. MR4632141

Aaron, C. and A. Cholaquidis (2020). On boundary detection. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 56*(3), 2028–2050. MR4116716

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall. MR0865647

Ambrosio, L., A. Colesanti, and E. Villa (2008). Outer Minkowski content for some classes of closed sets. *Mathematische Annalen 342*(4), 727–748. MR2443761

Balasubramanian, V., S.-S. Ho, and V. Vovk (2014). *Conformal Prediction for Reliable Machine Learning: Theory, Adaptations and Applications*. Newnes.

Berenfeld, C. and M. Hoffmann (2021). Density estimation on an unknown submanifold. *Electronic Journal of Statistics 15*(1), 2179–2223. MR4255312

Berry, T. and T. Sauer (2017). Density estimation on manifolds with boundary. *Computational Statistics & Data Analysis 107*, 1–17. MR3575055

Best, M. J. (2010). *Portfolio Optimization*. CRC Press. MR2722272

Bouzebda, S. and N. Taachouche (2024a). Oracle inequalities and upper bounds for kernel conditional u-statistics estimators on manifolds and more general metric spaces associated with operators. *Stochastics: An International Journal of Probability and Stochastic Processes 96*(8), 2135–2198. MR4843228

Bouzebda, S. and N. Taachouche (2024b). Rates of the strong uniform consistency with rates for conditional u-statistics estimators with general kernels on manifolds. *Mathematical Methods of Statistics 33*(2), 95–153. MR4773048

Chatterjee, S. (2021). A new coefficient of correlation. *Journal of the American Statistical Association 116*(536), 2009–2022. MR4353729

Cholaquidis, A., R. Fraiman, F. Gamboa, and L. Moreno (2023). Weighted lens depth: some applications to supervised classification. *Canadian Journal of Statistics 51*(2), 652–673. MR4595245

Cholaquidis, A., R. Fraiman, and L. Moreno (2022). Level set and density estimation on manifolds. *Journal of Multivariate Analysis 189*, 104925. MR4384138

Cholaquidis, A., R. Fraiman, and L. Moreno (2023). Level sets of depth measures in abstract spaces. *TEST*, 1–16. MR4656906

Cleanthous, G., A. G. Georgiadis, G. Kerkyacharian, P. Petrushev, and D. Picard (2020). Kernel and wavelet density estimators on manifolds and more general metric spaces. *Bernoulli 26*(3), 1832–1862. MR4091093

Cleanthous, G., A. G. Georgiadis, and E. Porcu (2022). Oracle inequalities and upper bounds for kernel density estimators on manifolds and more general metric spaces. *Journal of Nonparametric Statistics 34*(4), 734–757. MR4512277

Dheur, V., M. Fontana, Y. Estievenart, N. Desobry, and S. B. Taieb (2025). A unified comparative study with generalized conformity scores for multi-output conformal regression. *arXiv preprint arXiv:2501.10533*.

Diquigiovanni, J., M. Fontana, and S. Vantini (2022). Conformal prediction bands for multivariate functional data. *Journal of Multivariate Analysis 189*, 104879. MR4384121

Farouki, R. T., T. N. T. Goodman, and T. Sauer (2003). Construction of orthogonal bases for polynomials in Bernstein form on triangular and simplex domains. *Computer Aided Geometric Design 20*(4), 209–230. MR1984668

Fong, E. and C. C. Holmes (2021). Conformal bayesian computation. *Advances in Neural Information Processing Systems 34*, 18268–18279.

Fontana, M., S. Vantini, M. Tavoni, and A. Gammerman (2020). A conformal approach for distribution-free prediction of functional data. In *Functional and High-Dimensional Statistics and Related Fields 5*, pp. 83–90. Springer. MR4365783

Fontana, M., G. Zeni, and S. Vantini (2023). Conformal prediction: A unified review of theory and new challenges. *Bernoulli 29*(1), 1–23. MR4497237

Ghosal, S. (2001). Convergence rates for density estimation with Bernstein polynomials. *The Annals of Statistics 29*(5), 1264–1280. MR1873330

Guigui, N., N. Miolane, X. Pennec, et al. (2023). Introduction to Riemannian geometry and geometric statistics: from basic theory to implementation with geomstats. *Foundations and Trends® in Machine Learning 16*(3), 329–493. MR4209509

Hong, Y., R. Kwitt, N. Singh, N. Vasconcelos, and M. Niethammer (2016). Parametric regression on the Grassmannian. *IEEE Transactions on Pattern Analysis and Machine Intelligence 38*(11), 2284–2297.

Huang, Y. and G. Kou (2014). A kernel entropy manifold learning approach for financial data analysis. *Decision Support Systems 64*, 31–42.

Izbicki, R. and A. B. Lee (2017). Converting high-dimensional regression to high-dimensional conditional density estimation. *Electronic Journal of Statistics 11*(2), 2800–2831. MR3679910

Izbicki, R., G. Shimizu, and R. B. Stern (2022). Cd-split and hpd-split: Efficient conformal regions in high dimensions. *Journal of Machine Learning Research 23*(87), 1–32. MR4576672

Jammalamadaka, S. R. and Y. R. Sarma (1988). A correlation coefficient for angular variables. *Statistical Theory and Data Analysis II*, 349–364. MR0999196

Kuleshov, A., A. Bernstein, and E. Burnaev (2018). Conformal prediction in manifold learning. In *Conformal and Probabilistic Prediction and Applications*, pp. 234–253. PMLR.

Lee, J. M. (2013). *Introduction to Smooth Manifolds* (2nd ed.), Volume 218 of *Graduate Texts in Mathematics*. Springer. MR2954043

Lei, J. (2014). Classification with confidence. *Biometrika 101*(4), 755–769. MR3286915

Lei, J., A. Rinaldo, and L. Wasserman (2015). A conformal prediction approach to explore functional data. *Annals of Mathematics and Artificial Intelligence 74*, 29–43. MR3353895

Lei, J. and L. Wasserman (2014). Distribution-free prediction bands for nonparametric regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71–96. MR3153934

Leisch, F., E. Dimitriadou, K. Hornik, D. Meyer, and A. Weingessel (2007). mlbench: Machine learning benchmark problems. https://CRAN.R-project.org/package=mlbench. R package version 2.1-3.

Liu, H. and L. Wasserman (2016). Conformal prediction for multi-class classification. *Journal of Machine Learning Research 17*, 1–28.

Lugosi, G. and M. Matabuena (2024). Uncertainty quantification in metric spaces. *arXiv preprint arXiv:2405.05110*.

Mardia, K. V. and P. E. Jupp (2000). *Directional Statistics*, Volume 2. Wiley. MR1828667

Pawlowsky-Glahn, V., J. J. Egozcue, and R. Tolosana-Delgado (2015). *Modeling and Analysis of Compositional Data*. John Wiley & Sons. MR3328965

Pennec, X., S. Sommer, and T. Fletcher (2019). *Riemannian Geometric Statistics in Medical Image Analysis*. Academic Press. MR4254211

Petersen, A. and H.-G. Müller (2019). Fréchet regression for random objects with Euclidean predictors. *arXiv preprint arXiv:1608.03012*. MR3909947

Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters—an excess mass approach. *The Annals of Statistics*, 855–881. MR1345204

Rigollet, P. and R. Vert (2009). Optimal rates for plug-in estimators of density level sets. *Bernoulli 14*, 1154–1178. MR2597587

Romano, Y., E. Patterson, and E. Candès (2019). Conformalized quantile regression. *Advances in Neural Information Processing Systems 32*.

Thäle, C. (2008). 50 years sets with positive reach—a survey. *Surveys in Mathematics and its Applications 3*, 123–165. MR2443192

Tsybakov, A. B. (1997). On nonparametric estimation of density level sets. *The Annals of Statistics 25*(3), 948–969. MR1447735

Vovk, V., A. Gammerman, and G. Shafer (1998). Algorithmic learning in a random world. *Machine Learning 30*(2–3), 119–138. MR2161220

Vovk, V., A. Gammerman, and G. Shafer (2023). *Algorithmic Learning in a Random World* (2nd ed.). Springer. MR2161220

Wäschle, K., M. Bobak, S. Pölsterl, S. Gehrmann, D. Dettmering, J. Hornegger, and A. Maier (2014). Conformal prediction for regression. *IEEE Transactions on Medical Imaging 33*(2), 407–418.

Wu, H.-T. and N. Wu (2022). Strong uniform consistency with rates for kernel density estimators with general kernels on manifolds. *Information and Inference: A Journal of the IMA 11*(2), 781–799. MR4474348