Proceedings of Machine Learning Research 266:1–21, 2025 Conformal and Probabilistic Prediction with Applications

Anytime-Valid Tests of Group Invariance through Conformal Prediction

Tyron Lardy T.D.LARDY@MATH.LEIDENUNIV.NL

Centrum Wiskunde & Informatica and Leiden University
Muriel F. Pérez-Ortiz M.F.PEREZ.ORTIZ@TUE.NL
Eindhoven University of Technology

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

Abstract

Many standard statistical hypothesis tests, including those for normality and exchangeability, can be reformulated as tests of invariance under a group of transformations. We develop anytime-valid tests of invariance under the action of general compact groups and show their optimality—in a specific logarithmic-growth sense—against certain alternatives. This is achieved by using the invariant structure of the problem to construct conformal test martingales, a class of objects associated to conformal prediction. We apply our methods to extend recent anytime-valid tests of independence, which leverage exchangeability, to work under general group invariances. Additionally, we show applications to testing for invariance under subgroups of rotations, which corresponds to testing the Gaussian-error assumptions in linear models.

Keywords: Anytime-Valid Testing, Hypothesis Testing, Group Invariance, Conformal Prediction

1. Introduction

Conventional statistical models, either explicitly or implicitly, introduce assumptions of distributional symmetry about the data. For example, any distribution under which data are independent and identically distributed is symmetric under permutations of the observations and any regression model with Gaussian errors is symmetric under certain rotations of the data. If these symmetries are actually present in the data, employing symmetric models is advantageous in tasks such as hypothesis testing (Lehmann et al., 2005; Pérez-Ortiz et al., 2022), estimation (Brown, 1966), and prediction (Cohen and Welling, 2016). On the other hand, if the symmetries are absent from the data, the use of a symmetric model may lead to poor performance in these same tasks. We address the problem of testing for the presence of a given symmetry in data.

The presence of a symmetry is formalized as a null hypothesis of distributional invariance under the action of an algebraic group. Perhaps the most prominent example is infinite exchangeability—the hypothesis that the distribution of any finite data sequence is invariant under the group of all permutations. The null hypothesis of exchangeability is tested by classic methods such as permutation tests (Fisher, 1936; Pitman, 1937) and rank tests (Sidak et al., 1999). In this context, tests of symmetry are popular because they offer a simple vehicle to model complex dependencies under minimal assumptions. Tests for more general symmetries have also been studied, including tests for rotational symmetry (Baringhaus, 1991), symmetries for data taking values on groups (Diaconis, 1988), among others symmetries (Lehmann and Stein, 1949; Chiu and Bloem-Reddy, 2023).

In the applications that interest us, data are analyzed as they are collected, and the decisions to either stop and reach a conclusion or to continue data collection may depend on what has been observed so far. Sequential tests that retain type-I error control under such flexible data collection schemes were proposed by Wald (1947) and have been studied under the name of power-one tests (Robbins and Siegmund, 1974; Lai, 1977), and, more recently, as anytime-valid tests (Ramdas et al., 2023). The most common way to build anytime-valid tests (and in some sense the only admissible one (Ramdas et al., 2020)) is to monitor the value of a test martingale (a nonnegative martingale with expected value equal to one under the null hypothesis)—see Section 2.

While anytime-valid tests of general symmetries have not received much attention, anytime-valid tests for infinite exchangeability are present in the literature. Sen and Ghosh (1973a,b, 1974) develop asymptotic approximations and law-of-the-iterated-logarithm-type inequalities for linear rank statistics that hold uniformly over the duration of the experiment. More recently, Ramdas et al. (2022) and Saha and Ramdas (2024) develop test martingales for the hypothesis of infinite exchangeability for binary and paired data, respectively, while Vovk et al. (2003, 2005) builds anytime-valid tests under no additional assumptions. The latter work relies heavily on tools from conformal prediction, and it is most conducive to generalization.

Conformal prediction, perhaps best known as a framework for uncertainty quantification for point predictors, can be used to produce test martingales—called conformal martingales in this context—under the hypothesis of exchangeability. Most crucially for our present purposes, Vovk et al. (2005) show that conformal martingales cannot only be built to test for infinite exchangeability, but also to test whether data are generated by a general class of sequential data-generating mechanisms, called online compression models (see Section 3). Given our current interest in developing tests for general symmetries, it is natural to ask whether distributionally symmetric models define online compression models. If that was the case, the conformal martingales built by Vovk et al. (2005) would automatically yield tests of distributional symmetry. Unfortunately, this is not true in general because not every group-invariant model defines a compression model (see Section 3).

Contribution The main contribution of this work is to show that the above difficulty can be circumvented: under natural conditions, a distributionally symmetric model does define an online compression model. Furthermore, we show that the resulting conformal martingales are optimal in a specific sense. Indeed, we show that the resulting martingales are likelihood ratios against implicit alternatives and prove that they are optimal—in a sense that is specified in Section 2—for testing against that particular alternative. We use these constructions to abstract and generalize existing tests of independence under the assumption of exchangeability (Henzi and Law, 2024) to tests of independence under general symmetries. Finally, we build tests for the Gaussian-error assumptions behind linear models by testing for invariance under subgroups of the orthogonal group.

Outline The rest of this document is organized as follows. Section 2 formally introduces the problem of anytime valid testing for distributional invariance, and the optimality criterion that we employ. Then, in Section 3, the connection between group-invariant distributions and online compression models is shown. This connection is used in Section 4 to construct test martingales against the hypothesis of distributional invariance; the optimality of this procedure is shown in Section 4.2. Section 5 shows applications to test the assumptions of linear models, testing sign-invariant exchangeability, and independence testing. Section 6 discusses a potential direction for future work.

2. Problem Statement, Test Martingales and Optimality

The hypothesis of symmetry Suppose that we observe data $X_1, X_2, \dots$ sequentially and that they take values in some topological space $\mathcal{X}$. In our examples, $\mathcal{X} = \mathbb{R}$. We assume neither that these observations are independent nor that they are identically distributed. Furthermore, for each $n = 1, 2, \dots$, we assume that G_n is a compact topological group (in the algebraic sense) that acts continuously on $\mathcal{X}^n$. A topological group is a group equipped with a topology under which the group operation, a function $G_n \times G_n \to G_n$, is continuous. A (left) group action is a map $\varphi : G_n \times \mathcal{X}^n \to \mathcal{X}^n$ that satisfies, for any $g, h \in G_n$ and $x^n \in \mathcal{X}^n$, that $\varphi(h, \varphi(g, x^n)) = \varphi(hg, x^n)$. When the action is clear from context, we write gx^n instead of $\varphi(g, x^n)$. In examples, G_n is a group of $n \times n$ matrices and acts on $\mathbb{R}^n$ by matrix multiplication. The null hypothesis $\mathcal{H}_0$ of invariance under the action $(G_n)_{n\in\mathbb{N}}$ is

$$\mathcal{H}_0 : gX^n \overset{\mathcal{D}}{=} X^n \quad \text{for all } g \in G_n \text{ and all } n \in \mathbb{N}, \tag{1}$$

where $X^n := (X_1, \dots, X_n)$ and "$\overset{\mathcal{D}}{=}$" signifies equality in distribution. At this level of generality, one can build pathological examples of (1) that cannot be tested; more structure is needed (see Section 3). The next example contains instances of (1).

Example 1 (Exchangeability, rotational symmetry, and compact matrix groups) *Let $S(n)$ be the group of all $n \times n$ permutation matrices (matrices with exactly one entry of 1 in each row and each column, and 0 in all other entries). For tests of infinite exchangeability, the null hypothesis (1) becomes testing whether $\pi(X_1, \dots X_n) \overset{\mathcal{D}}{=} (X_1, \dots X_n)$ for all $\pi \in S(n)$ and $n \in \mathbb{N}$. Similarly, when testing for invariance under rotations of the data, the relevant group is $G_n = \mathrm{O}(n)$, the orthogonal group— all $n \times n$ matrices O such that $O^T O = I$. See Section 5.2 for details. The action of permutation and orthogonal groups are special examples of the actions of classic compact matrix groups on $\mathbb{R}^n$ (Meckes, 2019). The problem resulting form replacing $S(n)$ in (1) with any such group is amenable to this approach.*

Anytime-valid tests We construct sequential tests for $\mathcal{H}_0$ as in (1) that are anytime-valid at some prescribed level $\alpha \in (0, 1)$. A sequential test is a sequence $(\varphi_n)_{n\in\mathbb{N}}$ of rejection rules $\varphi_n : \mathcal{X}^n \to \{0, 1\}$ and we say that it is anytime valid for $\mathcal{H}_0$ at level α if

$$Q(\exists n \in \mathbb{N} : \varphi_n = 1) \le \alpha \text{ for any } Q \in \mathcal{H}_0.$$

Notice that this is a type-I error guarantee that is valid uniformly over all sample sizes: the probability that the null hypothesis is ever rejected by $(\varphi_n)_{n\in\mathbb{N}}$ is controlled by α. The most common construction, and in some sense the only admissible one (Ramdas et al., 2020), for anytime-valid tests uses test martingales (Ramdas et al., 2020; Grünwald et al., 2024) and minima thereof, e-processes—see Ramdas et al. (2020) for a comprehensive overview. We now define them.

Test martingales A sequence of statistics of the data is a test martingale if it is nonnegative, starts at one, and is a martingale under every element of $\mathcal{H}_0$. Formally, let $\mathbb{G} = (\mathcal{G}_n)_{n\in\mathbb{N}}$ be a filtration of σ-algebras such that $\mathcal{G}_n \subseteq \sigma(X^n)$, where $\sigma(X^n)$ denotes the σ-algebra induced by X^n. A sequence of statistics $(M_n)_{n\in\mathbb{N}}$ that is adapted to $\mathbb{G}$ is a test martingale for $\mathcal{H}_0$ with respect to $\mathbb{G}$ if $\mathbf{E}_Q[M_n \mid \mathcal{G}_{n-1}] = M_{n-1}$ for all $Q \in \mathcal{H}_0$ and $M_0 = 1$. Under $\mathcal{H}_0$, test martingales take large values with small probability. This is quantified by Ville's inequality (Ville, 1939), which shows that $\varphi_n = \mathbf{1}\{M_n \geq 1/\alpha\}$ is anytime valid.

Lemma 1 (Ville's inequality) *Let $(M_n)_{n\in\mathbb{N}}$ be a test martingale with respect to some filtration $(\mathcal{G}_n)_{n\in\mathbb{N}}$ under all elements of $\mathcal{H}_0$, then $\sup_{Q\in\mathcal{H}_0} Q(\exists n \in \mathbb{N} : M_n \geq 1/\alpha) \leq \alpha$.*

Proof Fix $Q \in \mathcal{H}_0$. Doob's optional stopping theorem states that $\mathbb{E}_Q[M_\tau] \leq 1$ for any stopping time τ that is adapted to $(\mathcal{G}_n)_{n\in\mathbb{N}}$ (Durrett, 2019, Theorem 5.7.6). Markov's inequality implies that $Q(M_\tau > \frac{1}{C}) \leq C$ for any $C > 0$. Applying this to the stopping time $\tau^* = \inf\{n \in \mathbb{N} : M_n \geq \frac{1}{\alpha}\}$ shows the result. ■

Test martingales that make use of external randomization will also prove useful; we will call them randomized test martingales. For randomized test martingales, we append an independent random number $\theta_n \sim \text{Uniform}([0,1])$ to each X_n, that is, we let $Y_n = (X_n, \theta_n)$ and consider test martingales that are functions of Y_n rather than X_n.

E-processes Test martingales are examples of e-processes (Ramdas et al., 2023). An e-process is any nonnegative stochastic process $(E_n)_{n\in\mathbb{N}}$ such that $\mathbb{E}_Q[E_\tau] \leq 1$ for all $Q \in \mathcal{H}_0$ and all stopping times τ. Given a e-process E_n, the test $\phi_n = \mathbf{1}\left\{E_n \geq \frac{1}{\alpha}\right\}$ is anytime valid. Relatedly, the product of e-processes based on independent data can be used in a valid test. Indeed, suppose two e-processes E and E' are used on independent data with different stopping times, τ_1 and τ_2, yielding E_{τ_1} and E'_{τ_2}. Then, their product satisfies $\mathbb{E}_Q[E_{\tau_1}E'_{\tau_2}] \leq 1$ for all $Q \in \mathcal{H}_0$ and $\mathbf{1}\left\{E_{\tau_1}E'_{\tau_2} \geq 1/\alpha\right\}$ as type-I error smaller than α by Markov's inequality. This property is known as safety under optional continuation, as E'_{τ_2} could be the result of continuing (or replicating) a first experiment that observed E_{τ_1}.

Log-optimality This type of evidence aggregation by multiplication of e-processes motivates a natural optimality criterion. Indeed, suppose we were to repeatedly run a single experiment, using a fixed e-process E and stopping time τ. If we measure the total evidence by the cumulative product of the individual e-processes, then the asymptotic growth rate of our evidence under true distribution P will be $\mathbb{E}_P[\log E_\tau]$. The current literature on anytime-valid focuses on this type of criterion, which can be traced back to Kelly betting (Kelly, 1956). Variants of this criterion have more recently been studied under several names (Koolen and Grünwald, 2022; Grünwald et al., 2024), but here we shall simply refer to maximizers of this criterion as "log-optimal".

3. Sequential Group Actions are Online Compression Models

The hypothesis in (1) can only be tested if the statements regarding group invariance for each $n \in \mathbb{N}$ are consistent with each other; without any further restrictions, invariance of the data at one time may contradict the invariance of the data at a later time. To avoid such situations, we assume action of the group sequence $(G_n)_{n\in\mathbb{N}}$ on the sample space is sequential, as defined next. After the statement of this definition, we discuss its meaning.

Definition 2 (Sequential group action) *We say that the action of the sequence of groups $(G_n)_{n\in\mathbb{N}}$ on $(\mathcal{X}^n)_{n\in\mathbb{N}}$ is sequential if the following conditions hold for all $n \in \mathbb{N}$.*

(i) There is an inclusion map $\imath_{n+1} : G_n \to G_{n+1}$ that is a continuous group isomorphism between G_n and its image, and the image of G_n under $\imath_{n+1}$ is closed in G_{n+1}.

(ii) For all $g_n \in G_n$ and all $x^{n+1} \in \mathcal{X}^{n+1}$, $\mathrm{proj}_{\mathcal{X}^n}(\imath_{n+1}(g_n)x^{n+1}) = g_n(\mathrm{proj}_{\mathcal{X}^n}(x^{n+1}))$, where $\mathrm{proj}_{\mathcal{X}^n}(x_1, \dots, x_n, x_{n+1}) = (x_1, \dots, x_n)$ is the canonical projection map.

(iii) Let $g_n \in G_n$ and $g_{n+1} \in G_{n+1}$. For $x^{n+1} = (x_1, \dots, x_{n+1}) \in \mathcal{X}^{n+1}$, denote $(x^{n+1})_{n+1} = x_{n+1}$. Then, $g_{n+1} = \imath_{n+1}(g_n)$ if and only if, for all $x^{n+1} \in \mathcal{X}^{n+1}$, $(g_{n+1}x^{n+1})_{n+1} = x_{n+1}$.

In Definition 2, Item (i) gives an ordering of the sequence of groups by inclusion, Item (ii) ensures that this inclusion does not change the action of the groups on past data, and Item (iii) implies that the groups do not act on future data. As a result, invariance of X^{n-1} under G_{n-1} is implied by invariance of X^n under G_n and the individual statements of invariance in (1) for each n do not contradict each other. The instances of (1) in Example 1 satisfy this assumption; the details for rotational symmetry are found in Section 5.2.

We now show that, under the assumption that the group action is sequential, the null hypothesis of invariance is an online compression model. When the data is generated by an online compression model, the techniques developed for conformal prediction can be used to construct a conformal (test) martingale, as we will discuss in Section 4. Vovk et al. define online compression models in abstract terms; we use a simplified definition here.

Definition 3 (Online compression model, Vovk et al. (2005)) *An online compression model on $\mathcal{X}$ is a 3-tuple of sequences $((\sigma_n)_{n\in\mathbb{N}}, (F_n)_{n\in\mathbb{N}}, (Q_n)_{n\in\mathbb{N}})$, where $(\sigma_n)_{n\in\mathbb{N}}$ is a sequence of statistics $\sigma_n = \sigma_n(X^n)$—we call σ_n a summary of X^n—; $(F_n)_{n\in\mathbb{N}}$ is a sequence of functions such that $F_n(\sigma_{n-1}, X_n) = \sigma_n$; and $(Q_n)_{n\in\mathbb{N}}$ is a sequence of conditional distributions for (σ_{n-1}, X_n) given σ_n.*

To show how sequential group invariance defines an online compression model, we first recall some group theory. First, the orbit G_nX^n of X^n under the action of G_n is the set of all values that are reached by the action of G_n on X^n; i.e., $G_nX^n = \{gX^n : g \in G_n\}$. We pick a single element of $\mathcal{X}^n$ in each orbit—an orbit representative—and consider the map $\gamma_n : \mathcal{X}^n \to \mathcal{X}^n$ that takes each X^n to its orbit representative. We call γ_n an orbit selector[1] (see Section 5 for examples). Furthermore, because G_n is a compact group, there exists a unique G_n-invariant probability distribution μ_n, called the Haar (probability) measure (Bourbaki and Berberian, 2004, Chapter VII). The Haar measure plays the role of the uniform probability distribution on compact groups. Finally, it is a fact that data are uniformly distributed on its orbit conditionally on the orbit where it lays; formally, $X^n \mid \gamma_n(X^n) \overset{\mathcal{D}}{=} U\gamma_n(X^n) \mid \gamma_n(X^n)$, where $U \sim \mu_n$ independently of X (Eaton, 1989, Theorem 4.4).

We now show that if a sequence of groups $(G_n)_{n\in\mathbb{N}}$ acts sequentially on the data, any $(G_n)_{n\in\mathbb{N}}$-invariant distribution defines an online compression model. To this end, it we need

1. we assume that it is measurable. Such measurable orbit selectors exist under weak regularity conditions (see Bondar, 1976, Theorem 2) that hold in all our examples.

to specify the three ingredients from Definition 3. First, we use the orbit representative as summary statistic; i.e., we use $\sigma_n = \gamma_n$. Secondly, the sequence of conditional distributions of (σ_{n-1}, X_{n-1}) given σ_n is uniform over the orbits as remarked earlier, which fixes Q_n. Lastly, the next proposition shows that, for sequential group actions, σ_n can be computed as a function of σ_{n-1} and X_n, which fixes $(F_n)_{n\in\mathbf{N}}$. The proof of this proposition can be found in Appendix A. This discussion and the following proposition prove that a sequential group-invariant model indeed defines an online compression model, as we state in Corollary 5.

Proposition 4 *If the action of $(G_n)_{n\in\mathbb{N}}$ on $(\mathcal{X}^n)_{n\in\mathbb{N}}$ is sequential, then there exists a sequence $(F_n)_{n\in\mathbb{N}}$ of measurable functions $F_n : \mathcal{X}^{n-1}\times\mathcal{X} \to \mathcal{X}^n$ such that $F_n(\gamma_{n-1}(X^{n-1}), X_n) = \gamma_n(X^n)$, and $F_n(\,\cdot\,, X_n)$ is a one-to-one function of $\gamma_{n-1}(X^{n-1})$.*

Corollary 5 *Assume that the action of $(G_n)_{n\in\mathbb{N}}$ on $(\mathcal{X}^n)_{n\in\mathbb{N}}$ is sequential, let $\widetilde{\mu}_n$ be the uniform distribution on G_nX^n induced by the Haar measure μ_n on G_n, and let $(F_n)_{n\in\mathbb{N}}$ be as guaranteed by Proposition 4. Then the tuple $((\gamma_n(X^n))_{n\in\mathbb{N}}, (F_n)_{n\in\mathbb{N}}, (\widetilde{\mu}_n)_{n\in\mathbb{N}})$ is an online compression model on $\mathcal{X}$.*

4. Testing Group Invariance With Conformal Martingales

We now construct test martingales for the null hypothesis of distributional symmetry in (1) under a sequential group action. To this end, the invariant structure of $\mathcal{H}_0$ is used in tandem with conformal prediction to build a sequence of independent random variables $(R_n)_{n\in\mathbb{N}}$ with the following three properties: (1) The sequence $(R_n)_{n\in\mathbb{N}}$ is adapted to the data sequence with external randomization $(X_n, \theta_n)_{n\in\mathbb{N}}$, that is, for each $n \in \mathbb{N}$, $R_n = R_n(X_n, \theta_n)$; (2) under any element of $\mathcal{H}_0$ from (1), $(R_n)_{n\in\mathbb{N}}$ is a sequence of independent and identically distributed Uniform$[0,1]$ random variables; and (3) the distribution of $(R_n)_{n\in\mathbb{N}}$ is not uniform when departures from symmetry are present in the data.

The construction of these random variables is the subject of Section 4.1—additional definitions are needed—and their optimality is the subject of Section 4.2. Example 2 shows testing exchangeability, also treated by Vovk et al. (2005) and Fedorova et al. (2012). They call the statistics $R_1, R_2, \ldots$ p-values owing to their uniformity. We opt against that terminology here because typically only small p-values are interpreted as evidence against the null hypothesis. In this context, it is any deviation from uniformity that we interpret as evidence against the null hypothesis. For reasons that will become apparent soon, we call $R_1, R_2, \ldots$ (smoothed) orbit ranks (see Definition 7).

With the sequence $(R_n)_{n\in\mathbb{N}}$ at hand, test martingales against distributional invariance are built by testing against the uniformity of $(R_n)_{n\in\mathbb{N}}$. Indeed, any time that $(f_n)_{n\in\mathbb{N}}$ is a sequence of functions $f_i : [0,1] \to \mathbb{R}$ such that $\int f_i(r)\mathrm{d}r = 1$—called calibrator by (Vovk and Wang, 2021)—, the process $(M_n)_{n\in\mathbb{N}}$ given by

$$M_n = \textstyle\prod_{i\leq n} f_i(R_i) \tag{2}$$

is a test martingale under $\mathcal{H}_0$ with respect to $\mathbb{F} = (\sigma(R^n))_{n\in\mathbb{N}}$, where $\sigma(R^n)$ is the σ-algebra generated by R^n. This follows because $\mathbf{E}_Q\left[M_n \mid \sigma(R^{n-1})\right] = M_{n-1}\int f_n(r)\mathrm{d}r = M_{n-1}$ by independence and uniformity. As we will see, the sequence $(f_n)_{n\in\mathbb{N}}$ can be taken to be any sequence of predictable estimators of the density of $R_1, R_2, \ldots$ (Fedorova et al., 2012), so that the test martingale is expected to grow whenever the distribution of the orbit ranks is not uniform; i.e., if the null hypothesis is violated (see Section 4.2).

Example 2 (Sequential Ranks) *Consider the case of testing exchangeability as discussed in Example 1. For each n, let $\widetilde{R}_n = \sum_{i \leq n} \mathbf{1}\{X_i \leq X_n\}$ be the rank of X_n among $X_1, \ldots, X_n$. The random variables $\widetilde{R}_1, \widetilde{R}_2, \ldots$ are called sequential ranks (Malov, 1996). It is a classic observation that each $\widetilde{R}_n$ is uniformly distributed on $\{1, \ldots, n\}$, and that $(\widetilde{R}_n)_{n\in\mathbb{N}}$ is an independent sequence (Rényi, 1962). After rescaling and adding external randomization, a sequence $(R_n)_{n\in\mathbb{N}}$ can be built from $(\widetilde{R}_n)_{n\in\mathbb{N}}$ such that $(R_n)_{n\in\mathbb{N}}$ satisfies items 1, 2 and 3 at the start of this section. Furthermore, if we denote the uniform measure on $S(n)$ by μ_n, then $(1/n)\widetilde{R}_n = \mu_n\{g : (gX_n)_n \leq X_n\}$. While this rewriting may seem esoteric at this point, it turns out to be the correct point of view for generalization.*

4.1. Conformal Prediction Under Invariance

In general, the statistics R_n will be designed to measure how strange the observations X^n are in contrast to what would be expected under distributional invariance. To this end, the values of X^n are compared to those in the orbit of X^n under the action of G_n. To measure this "strangeness", we adapt the conformity measures introduced by Vovk et al. (2005).

Definition 6 (Conformity measure of invariance) *The function $A : \mathcal{X} \times \bigcup_{n=1}^{\infty} \mathcal{X}^n \to \mathbb{R}$ is a conformity measure of invariance if the following holds: if there are $X^n, X'^n \in \mathcal{X}^n$, such that $A(X_i, \gamma_n(X^n)) = A(X'_i, \gamma_n(X'^n))$ for all $i \in \{1, \ldots, n\}$, then, for all $g \in G_n$, we also have that $A((gX^n)_i, \gamma_n(X^n)) = A((gX'^n)_i, \gamma_n(X'^n))$ for all $i \in \{1, \ldots, n\}$.*

The group-related condition on A in Definition 6 is an addition to that of Vovk et al. (2005); it ensures that the action of G_n on $\mathcal{X}^n$ induces an action on the conformity measure. When A is properly chosen, $A(X_n, \gamma_n(X^n))$ indicates how similar X_n is to the other values in its orbit. For this reason, the statistic $\alpha_n = A(X_n, \gamma_n(X^n))$ is called a conformity score. We now give some examples. If $\mathcal{X}^n \subseteq \mathbb{R}^n$ then $A(X_n, \gamma_n(X^n)) = X_n$ is a conformity measure of invariance—as in Example 2. When each observation X_n takes values in $\mathbb{R}^d$, more sofisticated measures are needed. In that case, a natural choice is $A(X_n, \gamma_n(X^n)) = \|X_n - \int_G g\gamma_n(X^n)\,\mathrm{d}\mu_n(g)\|^{-1}$, which takes large values whenever X_n is close to the orbit average. For a more involved example, consider outcome-covariate observations $X_i = (Y_i, Z_i)$ for outcomes $Y_i \in \mathbb{R}^d$ and a covariates $Z_i \in \mathbb{R}$. Consider $A(X_i, \gamma_n(X^n)) = \|Y_i - \hat{Y}_i\|^{-1}$, where $\hat{Y}_i$ is a prediction, for example, using a regression method trained on the orbit of X_i. Intuitively, if the label is very close to the prediction made using all values in its orbit, then X_i is typical in the orbit. For a detailed discussion on conformity measures, see Fontana et al. (2023). However, given the statistical nature of the examples that we treat, the simple case that $X_n \in \mathbb{R}$ and $A(X_n, \gamma_n(X^n)) = X_n$ will be the most prominent here. Design of conformity scores for testing general invariances remains an interesting direction.

Since the scale of the conformity scores is arbitrary—they can be scaled at will—, only comparisons between them are meaningful. Therefore, as an extension of Example 2, we rank the observed value of the conformity score α_n among all its possible values on the orbit of the data. Indeed, as discussed in Section 3, the condition distribution of X^n given $\gamma(X^n)$ is uniform on its orbit. We use this to obtain the distribution of the conformity scores and motivates the (smoothed) orbit ranks $(R_n)_{n\in\mathbb{N}}$ in the next definition.

Definition 7 (Smoothed Orbit Ranks) *Fix $n \in \mathbb{N}$, let A be a conformity measure, and let $\alpha_n = A(X_n, \gamma_n(X^n))$ be the associated conformity score. We call R_n, defined by*

$$R_n = \mu_n(\{g : A((gX^n)_n, \gamma_n(X^n))_n < \alpha_n\}) + \theta_n \mu_n(\{g : A((gX^n)_n, \gamma_n(X^n)) = \alpha_n\}), \quad (3)$$

a (smoothed) orbit rank, where μ_n is the Haar probability measure on G_n and $\theta_n \sim \mathrm{Uniform}[0,1]$ is independent of the data X^n.

When the group G_n is finite of size k and $A(X_i, \gamma_n(X^n)) = X_i$, μ_n is the discrete uniform distribution on G_n and $R_n = \frac{1}{k}\#\{g \in G_n : (gX^n)_n < X_n\} + \frac{\theta_n}{k}\#\{g \in G_n : (gX^n)_n = X_n\}$.

Vovk et al. (2005, Theorem 11.2) show that, if the data are generated by an online compression model, and $\theta_1, \theta_2, \ldots$ are independent, then $R_1, R_2, \ldots$ are also independent. Since Corollary 5 shows that a sequential group invariance structure defines an online compression model, it follows that the smoothed orbit ranks form an i.i.d. uniform sequence under the null hypothesis. This is stated in the next theorem, for which we provide a direct proof in Appendix A for completeness .

Theorem 8 *Suppose that the action of $(G_n)_{n\in\mathbb{N}}$ on $(\mathcal{X}^n)_{n\in\mathbb{N}}$ is sequential, that $(X_n)_{n\in\mathbb{N}}$ is generated by an element of $\mathcal{H}_0$, and that $\theta_1, \theta_2, \ldots$ are independent. Then $R^n \perp \gamma_n(X^n)$ for each n and $(R_n)_{n\in\mathbb{N}}$ is a sequence of i.i.d. $\mathrm{Uniform}[0,1]$ random variables.*

4.2. Optimality

We now show that any martingale based on the smoothed orbit ranks as in (2) is a likelihood ratio process and that it is log-optimal against the implicit alternative for which it is built. To this end, let P be a distribution such that for all n, conditionally on R^{n-1}, R_n has density f_n (with respect to the Lebesgue measure). Technically, the conditional density of R_n is not defined only by P, but also by the external randomization. To make this explicit, we write $\widetilde{P}$ to denote P with external randomization, that is, $\widetilde{P} = P \times \mathcal{U}^\infty$, where $\mathcal{U}^\infty$ is the uniform distribution on $[0,1]^\infty$. Analogously, for each $Q \in \mathcal{H}_0$, define $\widetilde{Q} = Q \times \mathcal{U}^\infty$.

The discussion below (2) shows that $M_n = \prod_{i\le n} f_i(R_i)$ is a test martingale. In fact, M_n is the likelihood ratio for the orbit ranks R^n between $\widetilde{P}$ and $\widetilde{Q}$, since the distribution of R^n under $\widetilde{Q}$ equals the uniform distribution for any $Q \in \mathcal{H}_0$ by Theorem 8. Surprisingly, if $\widetilde{P}$ is such that $R_n \perp \gamma_n(X^n)$, then M_n is also the likelihood ratio for the full data X^n between P and an appropriately chosen distribution $Q^* \in \mathcal{H}_0$, as shown in the following proposition. We assume throughout that the action of $(G_n)_{n\in\mathbb{N}}$ on $(\mathcal{X}^n)_{n\in\mathbb{N}}$ is sequential.

Proposition 9 *Suppose that $A(\,\cdot\,, \gamma_n(X^n))$ is a one-to-one function for each $n \in \mathbb{N}$, suppose that P is any distribution under which $R_n \perp \gamma_n(X^n)$ for each n, and let f_i denote the conditional density of R_i given R^{i-1} under P. Let $M_n = \prod_{i\le n} f_i(R_i)$. Then, for $Q \in \mathcal{H}_0$,*

$$\widetilde{Q}\left(M_n = \tfrac{\mathrm{d}P}{\mathrm{d}Q^*}(X^n)\right) = 1, \quad (4)$$

where Q^* denotes the distribution under which the marginal of $\gamma_n(X^n)$ coincides with that under P, and such that $X^n \mid \gamma_n(X^n) \overset{\mathcal{D}}{=} U\gamma_n(X^n) \mid \gamma_n(X^n)$, where $U \sim \mu_n$ independently from $\gamma_n(X^n)$.

The distribution Q^* can be thought of as a symmetrization of P, since the marginal of $\gamma_n(X^n)$ is the same, but the distribution conditional on $\gamma_n(X^n)$ is defined by symmetry. Proposition 9 therefore shows that, if the orbit ranks are independent of the orbit selectors under P, then $(M_n)_{n\in\mathbb{N}}$ is the likelihood ratio process between P and a symmetrization thereof. The next theorem uses this representation to show the log-optimality of $(M_n)_{n\in\mathbb{N}}$. Its proof follows that of Theorem 12 of Koolen and Grünwald (2022).

Theorem 10 *Assume that $A(\,\cdot\,, \gamma_n(X^n))$ is one-to-one for all $n \in \mathbb{N}$ and let P be such that, under P, the distribution of $X^n \mid \gamma_n(X^n)$ is absolutely continuous with respect to the uniform distribution. Denote f_i for the density of $R_i \mid R^{i-1}$ under $\widetilde{P}$ and let $M_n = \prod_{i\leq n} f_i(R_i)$. Let τ be any stopping time and $(E_n)_{n\in\mathbb{N}}$ any e-process for $\mathcal{H}_0$, both with respect to $\mathbb{F}$—the filtration generated by the smoothed ranks. Then it holds that*

$$\mathbf{E}_{\widetilde{P}}\left[\ln M_\tau\right] \geq \mathbf{E}_{\widetilde{P}}\left[\ln E_\tau\right]. \tag{5}$$

Moreover, if $\widetilde{P}$ is such that $R^n \perp \gamma_n(X^n)$ for all n, then for any e-process $E^{\text{full-data}}$ for $\mathcal{H}_0$ w.r.t. $(\sigma(X^n, \theta^n))_{n\in\mathbb{N}}$—the full-data filtration—, it also holds that

$$\mathbf{E}_{\widetilde{P}}\left[\ln M_\tau\right] \geq \mathbf{E}_{\widetilde{P}}\left[\ln E_\tau^{\text{full-data}}\right]. \tag{6}$$

The first part of Theorem 10, (5), establishes that, under some assumptions on P, $(M_n)_{n\in\mathbb{N}}$ is log-optimal for testing group invariance among all e-processes defined only on the orbit ranks. The second part of Theorem 10, (6), states that if the orbit ranks are also independent of the orbit selector under P, then $(M_n)_{n\in\mathbb{N}}$ is log-optimal for testing group invariance among all e-processes defined on the full data(!). The additional assumption of independence between R^n and $\gamma_n(X^n)$ is necessary for (6) to hold. Indeed, without independence and for $\tau = n$, the log-optimal statistic is the likelihood ratio $S_n = \prod_{i=1}^n f_n(R_1, \ldots, R_n \mid \gamma_n(X^n))$ (see also Grünwald et al., 2024; Koning, 2023). However, the sequence $(S_n)_{n\in\mathbb{N}}$ does not always define an e-process, and it is not clear how to use it for anytime-valid testing. Tests based on the sequential ranks circumvents this issue.

The optimality of M_n in Theorem 10 requires oracle knowledge of the true distributions $f_1, f_2, \ldots$, which are unknown in practice. To counter this, past data can be used to estimate the true density (Vovk et al., 2005; Fedorova et al., 2012). More precisely, for each n, let $\hat{f}_n$ be an estimator of f_n based on R^{n-1}, and consider the martingale defined by $\prod_{i=1}^n \hat{f}_i(R_i)$. In the case that there exists a density f such that $f_i \equiv f$ for all i, i.e. data are i.i.d. under P, there is limited loss asymptotically if $\hat{f}_i$ is a good estimator of f. In order to judge if an estimator is good for the task at hand, consider the difference in expected growth per outcome for fixed n; i.e., $\frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\widetilde{P}}[\log f(R_i) - \log \hat{f}_i(R_i)] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\widetilde{P}}[\mathrm{KL}(f\|\hat{f}_i)]$, where $\mathrm{KL}(f\|g) = \int_0^1 f(r)\log(f(r)/g(r))\mathrm{d}r$ is the Kullback-Leibler divergence whenever f is absolutely continuous with respect to g, and the expectation $\mathbb{E}_{\widetilde{P}}$ is over past data (on which $\hat{f}_i$ depends). If $\frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\widetilde{P}}[\mathrm{KL}(f\|\hat{f}_i)] \to 0$ as n grows large, the expected growth per outcome converges to that of the log-optimal test martingale. Under weak assumptions, specialized algorithms exist with this guarantee (Haussler and Opper, 1997; Cesa-Bianchi and Lugosi, 2001; Grünwald and Mehta, 2019).

5. Applications and Extension

We discuss applications and an extension of the theory developed in the previous sections. In Table 1, we give a summary of all examples.

5.1. Sign-Invariant Exchangeability: A Simulation

In this subsection, we consider testing for sign-invariant exchangeability (Berman, 1965; Fraiman et al., 2024) with the purpose of illustrating our method on a concrete example, and show its performance through numeric simulation. Real-valued data are sign-invariant if $(X_1, \ldots, X_n) \overset{\mathcal{D}}{=} (\epsilon_1 X_1, \ldots, \epsilon_n X_n)$ for all signs $(\epsilon_1, \ldots, \epsilon_n) \in \{-1, 1\}^n$. We consider $\{-1, 1\}^n$ as a group with componentwise multiplication as operation. Data are sign-invariant exchangeable if they are both sign-invariant and exchangeable; i.e., if for all signs $(\epsilon_1, \ldots, \epsilon_n) \in \{-1, 1\}^n$ and all permutations $\pi \in S(n)$ we have $(X_1, \ldots, X_n) \overset{\mathcal{D}}{=} (\epsilon_1 X_{\pi(1)}, \ldots, \epsilon_n X_{\pi(n)})$. The null hypothesis of sign-invariant exchangeability is equivalent to distributional invariance of X^n under the action of $G_n = \{-1, 1\}^n \times S(n)$. The orbit of X^n under G_n is $\{(\epsilon_1 X_{\pi(1)}, \ldots, \epsilon_n X_{\pi(n)}) : (\epsilon_1, \ldots, \epsilon_n) \in \{-1, 1\}^n, \pi \in S(n)\}$ and the orbit selector is $\mathrm{sort}(|X_i| : i \leq n)$, where $\mathrm{sort} : \mathbb{R}^n \to \mathbb{R}^n$ sorts its argument. Since G_n is finite, the Haar measure is the discrete uniform distribution. We can take $A(X_n, \gamma_n(X^n)) = X_n$. Let $X'_1, \ldots, X'_{2n}$ be given, for $i = 1, \ldots, n$, by $X'_i = X_i$ and $X'_{n+i} = -X_i$. The smoothed orbit rank in (3) becomes

$$R_n = \tfrac{1}{2n} \#\{i \leq 2n : X'_i < X_n\} + \tfrac{\theta_n}{2n} \#\{i \leq 2n : X'_i = X_n\}. \tag{7}$$

Standard density estimation algorithms can used to estimate their density. Following Section 4, for each n we use $R_1, \ldots, R_{n-1}$ to build an estimate $\hat{f}_n$ of the density and use $M_n = \prod_{i \leq n} \hat{f}_i(R_i)$ as a test martingale.

We investigate numerically how martingales obtained in this manner behave through simulations with a specific choice of density estimators (see Appendix C for more details). For the sake of comparison, we also include the conformal martingale that would be obtained if testing for either exchangeability alone or for sign-invariance alone. The results are shown in Figure 1. If data are sign-invariant exchangeable—i.i.d. Rademacher or i.i.d. Normal$(0, 1)$ in our experiments—, the conformal martingales do not take large values (not shown). Under the alternative, the statistic M_n is no longer a martingale, and it does grow. However, the methods that test for only one of the two symmetries (either exchangeability or sign invariance separately) do not detect alternatives for which that particular symmetry is not violated, but the other is (see Figure 1). On the other hand, the conformal martingale based on R_n as described in (7) detects all of the alternatives. In fact, for the alternative where each $X_i \in \{-1, 1\}$ and $X_i = 1$ with probability $p_i = 1 - 1/i$ and $X_i = -1$ with probability $1 - p_i$ independently, the corresponding test martingale is even log-optimal among all e-processes. This is due to the fact that, regardless of the observed data, the orbit of X_i is always the set $\{-1, 1\}$. Therefore, the orbit selector can be chosen to be $\gamma_n(X^n) = (1, \ldots, 1)$ independently of the data, such that $R_n \perp \gamma_n(X^n)$. The log-optimality then follows from Theorem 10.

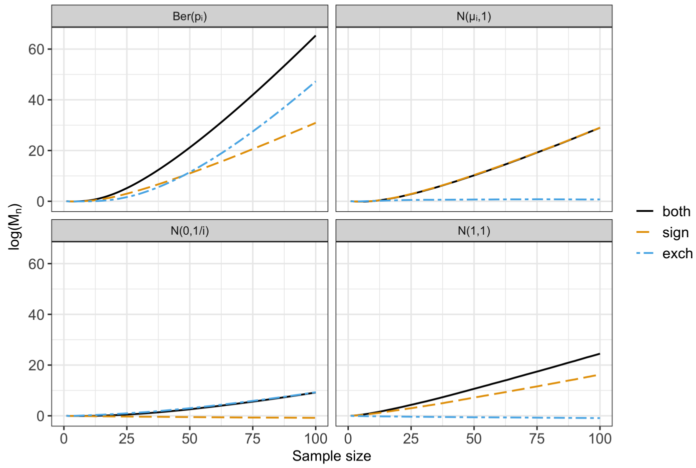

Figure 1: Logarithmic conformal martingale against sample size for three different methods: testing for both sign-invariance and exchangeability, or either of them. Upper left corner: $X_i = 1$ w.p. $p_i = 1 - 1/i$ and $X_i = 1$ with probability $1 - p_i$. Upper right corner: $X_i \sim N(i/10, 1)$. Lower left corner: $X_i \sim N(0, 1/i)$. Lower right corner: $X_i \sim N(1, 1)$. A test built against sign-invariance and exchangeability can detect the absence of either of those two invariances while a test that is built to detect only one of them cannot (see Section 5.1). Data are independent under all considered models. The results were averaged over 500 repetitions.

5.2. The Orthogonal Group and Linear Models

Consider testing whether the data are drawn from a spherically symmetric distribution; i.e., $\mathcal{X} = \mathbb{R}$ and $G_n = \mathrm{O}(n)$, where $\mathrm{O}(n)$ is the orthogonal group in dimension n. Testing for spherical symmetry is equivalent to testing whether the data are generated by a zero-mean Gaussian distribution. Indeed, any distribution on $\mathbb{R}^\infty$ for which the marginal of the first n coordinates is spherically symmetric for any n, is a mixture of i.i.d. zero-mean Gaussian distributions (Bernardo and Smith, 2009, Proposition 4.4). It follows that any process that is a supermartingale under all zero-mean Gaussian distributions is also a supermartingale under spherical symmetry and vice-versa. This implies that the two hypotheses are equivalent. We show how this fits in our setting, and defer the application to regression to Appendix B.

We now check that testing spherical symmetry fits in our setting, i.e., that Definition 2 is fulfilled. Consider the inclusion of $\mathrm{O}(n)$ in $\mathrm{O}(n+1)$ given by $\imath_{n+1}(O_n) = \begin{pmatrix} O_n & 0 \\ 0 & 1 \end{pmatrix}$ for each $O_n \in \mathrm{O}(n)$. Using the canonical projections in $\mathbb{R}^n$, Definition 2 is readily checked. Since the data are real, we can consider the measure of conformity $A(X_n, \gamma_n(X^n)) = X_n$. An orbit selector is given by $\gamma_n(X^n) = \|X^n\| e_1$, where e_1 is the unit vector $e_1 = (1, 0, \ldots, 0)$. For simplicity, we assume that the distribution of X^n has a density with respect to the Lebesgue measure for each n, so that $R_n = \mu_n(\{O_n \in \mathrm{O}(n) : (O_n X^n)_n < X_n\})$—no external randomization is needed. Rather than thinking of μ_n as a measure on $O(n)$, one can think of it as the uniform measure on $S^{n-1}(\|X^n\|)$. This way, R_n can be recognized to be the relative surface area of the hyper-spherical cap with co-latitude angle $\varphi_n = \pi - \cos^{-1}(X_n/\|X^n\|)$. Li (2010) shows that this area is given by

$$R_n = \begin{cases} 1 - \frac{1}{2} I_{\sin^2(\pi - \varphi_n)}\left(\frac{n-1}{2}, \frac{1}{2}\right) & \text{if } \varphi_n > \frac{\pi}{2}, \\ \frac{1}{2} I_{\sin^2(\varphi_n)}\left(\frac{n-1}{2}, \frac{1}{2}\right) & \text{else}, \end{cases} \tag{8}$$

where $I_x(a,b)$ is the regularized beta function, $I_x(a,b) = \frac{B(x,a,b)}{B(1,a,b)}$ for $B(x,a,b) = \int_0^x t^{a-1}(1-t)^{b-1}$ and $0 \leq x \leq 1$.

Note that $\varphi_n > \frac{\pi}{2}$ if and only if $X_n > 0$ and that $\sin^2(\varphi_n) = 1 - \frac{X_n^2}{\|X^n\|^2}$, so that (8) equals the CDF of the t-distribution with $n-1$ degrees of freedom evaluated in $t = \sqrt{n-1}X_n/\|X^{n-1}\|$. If $X^n \sim \mathcal{N}(0, \sigma^2 I_n)$, then t is the ratio of a normally distributed random variable and an independent chi-squared-distributed random variable. Therefore, t has a t-distribution with $n-1$ degrees of freedom. The test thus obtained is a type of sequential t-test. To the best of our knowledge, this test has not been considered previously.

We show extensions to symmetry around a point other than the origin and more general symmetries relevant to regression in Appendix B.

5.3. Modification for Independence Testing

We now propose a minor modification of the conformal martingales from the previous section that can be used to test for independence. Formally, fix $K \in \mathbb{N}$ and suppose that at each time point $n \in \mathbb{N}$, a K-dimensional vector $X_n = (X_{1,n}, \dots, X_{K,n}) \in \mathcal{X}^K$ is observed. We are interested in testing the null hypothesis that states that: (1) for each $k = 1, \dots, K$ and each n the vectors $(X_{k,1}, \dots, X_{k,n})$ are G_n-invariant, and (2) $(X_{k,1}, \dots, X_{k,n}) \perp (X_{k',1}, \dots, X_{k',n})$ for all $k \neq k' \in \{1, \dots, K\}$. Under this hypothesis, the data is invariant under the sequential action of $(\widetilde{G}_n)_{n \in \mathbb{N}}$ given by $\widetilde{G}_n = G_n^K$, acting on $\mathcal{X}^{K \times n}$ rowwise. That is, the first copy of the group acts on $(X_{1,1}, \dots, X_{1,n})$, the second on $(X_{2,2}, \dots, X_{2,n})$, etc. This action is sequential anytime that the action of $(G_n)_{n \in \mathbb{N}}$ is sequential on each of the K data streams.

A first idea to test for invariance under $\widetilde{G}_n$ is to create K test martingales and combine them through multiplication. More specifically, we can treat each $(X_{k,n})_{n \in \mathbb{N}}$, $k \in \{1, \dots, K\}$ as a separate data stream and compute, for a choice of conformity measyre, the corresponding statistics in (3), leading to K sequences of uniformly distributed random variables $(R_{k,n})_{n \in \mathbb{N}}$. If, for all $n \in \mathbb{N}$ and $k \in \{1, \dots, K\}$, $f_{k,n}$ is a density on $[0,1]$ then, by independence, the sequence $(M'_n)_{n \in \mathbb{N}}$ defined by $M'_n = \prod_{i=1}^n \prod_{k=1}^K f_{k,i}(R_{k,i})$ is a martingale under the null hypothesis. Unfortunately, this martingale would not be able to detect alternatives under which the marginals are group invariant, but not independent. Indeed, it only uses that the marginals are uniform under the null, while in fact a stronger claim is true: for each n, the joint distribution of $R_{k,n}$, $k \in \{1, \dots, K\}$, is uniform on $[0,1]^K$. One must instead choose any sequence of joint density (estimators) $f_1, f_2, \dots$ on $[0,1]^K$ and use a test martingale $M_n = \prod_{i=1}^n f_i(R_{1,i}, \dots, R_{K,i})$.

When $K = 2$ and $G_n = S(n)$, this procedure was proposed by Henzi and Law (2024). They show that if f_n is a particular histogram density estimator, it is possible to detect departures from independence consistently under the stronger assumption that data are i.i.d. One of their key insights is that independence of the data streams not only implies joint uniformity of the sequential ranks in their setting, but that independence and joint uniformity are actually equivalent. This equivalence breaks down without the assumption that $X_{k,1}, X_{k,2}\dots$ are i.i.d. for all k. Finding conditions under which the independence of the streams and the joint uniformity of the rank distributions are equivalent in a more general setting is an interesting avenue of research.

G_n	γ_n	R_n
$S(n)$	$\mathrm{sort}(X^n)$	$\frac{1}{n}(1-\theta+\sum_{i\leq n}\mathbf{1}\{X_i \leq X_n\})$
$O(n)$	$\lVert X^n\rVert e_1$	see (8)
$\{O \in O(n) : O\mathbf{1}_n = \mathbf{1}_n\}$	$\lVert X^n_{\perp \mathbf{1}_n}\rVert e_1'$, $e_1' \in S^{n-1}(1)$	see (8) and (9)
$\{-1,1\}^n \times S(n)$	$\mathrm{sort}(\lvert X_i\rvert : i \leq n)$	see (7)

Table 1: Summary of examples discussed. They correspond, in descending order, to testing for exchangeability without tied observations, in Example 1; testing rotations, in Example 1; testing for sphericity, in Section 5.2; testing sphericity of residual errors in linear regression with nonzero mean, in Appendix B; and testing for sign-invariant exchangeability, in Section 5.1. In all cases, data is given by $X_n \in \mathbb{R}$ and $A(X_n, \gamma_n(X^n)) = X_n$.

6. Conclusion and Discussion

This article shows how to apply the theory of conformal prediction to test for symmetries in data sequentially. It was shown that any distribution on data sequences that is invariant under a compact group can be used to build an online compression model (Section 3). These compression models are used to build conformal martingales (Section 4). Applications to regression, sign-invariant exchangeability and rotations are shown in Section 5.

Here we discuss three topics. First, the relationship to noninvariant conformal martingales. Second, whether smoothing is necessary when defining orbit ranks. Third, the design of conformity scores.

Noninvariant Conformal Martingales Not all online compression models correspond to a compact-group invariant null hypothesis. An interesting example is when the data are i.i.d. and exponentially distributed. This distribution is invariant under reflections in any 45° line (not necessarily through the origin), but these reflections do not define a compact group and therefore do not fit the setting discussed in this article. Nevertheless, the sum of data points is a sufficient statistic for the data, so this model can still be seen as an online compression model with the sum being the summary. More work is needed to find out whether conformal martingales are log-optimal against certain alternatives in such settings.

The Need for Smoothing When the conformity measure $\alpha^n(X^n)$ conditionally on the orbit selector $\gamma_n(X^n)$ has a continuous distribution, no smoothing is necessary in (7). This is the case for the rotations discussed in Section 5.2. In other scenarios, smoothing can be avoided as well. Indeed, one can always define nonsmoothed orbit ranks $\widetilde{R}_n := \mu_n(\{g \in G_n : (g\alpha^n)_n \leq \alpha_n\})$—this is setting θ_n to 0 in Definition 7. We have that $\widetilde{R}_n \leq R_n$. If the densities $f_1, f_2, \dots,$ on $[0,1]$ are nondecreasing—in the sense that $u \mapsto f_i(u)$ is nondecreasing—, the process $\widetilde{M}_n := \prod_{i=1}^n f_i(\widetilde{R}_i)$ is bounded from above by the conformal martingale $M_n = \prod_{i=1}^n f_i(R_i)$. Such a choice of increasing f_i is natural when large R_i

(or $\widetilde{R}_i$) is associated with departures from the null hypothesis. Therefore, a test based on $\widetilde{M}_n$ inherits the anytime-valid type-I error guarantees from M_n because $\widetilde{M}_n \leq M_n$ Vovk et al. (2003). However, the process $\widetilde{M}_n$ may not be a martingale itself. Instead, a test martingale can sometimes directly be associated to $\widetilde{R}_n$. For instance, in the setting of Example 2 (testing exchangeability), the null distribution of $\widetilde{R}_n$ is known—it is uniformly distributed on $\{1, \ldots, n\}$. Therefore, we can normalize each multiplicative increment by its null expectation. Even more, there are parametric alternatives under which the exact distributions of the nonsmoothed ranks are closed form. This is the case for Lehmann alternatives where each X_i is assumed to be sampled from some continuous distribution with c.d.f. $F_i(x) = 1 - (1 - F_0(x))^{\theta_i}$ for some θ_i. Theorem 7.a.1 of Savage (1956) gives the distribution of $\widetilde{R}_i$. The exact likelihood ratio process of $\widetilde{R}_i$ is log-optimal and avoids external randomization.

Conformity scores of invariance In this work, the main examples dealt with real observations. This is because of the initial interest of the authors in statistical applications. In these cases, the conformity score $A(X_n, \gamma_n(X^n))$ can be taken to be $A(X_n, \gamma_n(X^n)) = X_n$, resulting in powerful tests against invariance (see Section 5.1). However, in more general situations where each observation is not real, as is the case in many current Machine Learning applications, this simple conformity score is not available. Designing and assessing conformity scores of invariance remains an interesting direction.

Acknowledgments We thank the attendees of the Seminar on Anytime-Valid Inference "e-readers" at *Centrum Wiskunde & Informatica* in Amsterdam for their input and valuable insights. In particular, we are grateful to Peter Grünwald for feedback on a first version of this article, and Nick Koning for fruitful discussions.

References

Ludwig Baringhaus. Testing for spherical symmetry of a multivariate distribution. *The Annals of Statistics*, pages 899–917, 1991.

Simeon M Berman. Sign-invariant random variables and stochastic processes with sign-invariant increments. *Transactions of the American Mathematical Society*, 119(2):216–243, 1965.

José M Bernardo and Adrian FM Smith. *Bayesian theory*, volume 405. John Wiley & Sons, 2009.

James V Bondar. Borel cross-sections and maximal invariants. *The Annals of Statistics*, pages 866–877, 1976.

N. Bourbaki and S.K. Berberian. *Integration II: Chapters 7–9*. Springer Berlin Heidelberg, 2004. ISBN 9783540205852.

Lawrence David Brown. On the admissibility of invariant estimators of one or more location parameters. *The Annals of Mathematical Statistics*, 37(5):1087–1136, 1966.

Nicolo Cesa-Bianchi and Gabor Lugosi. Worst-case bounds for the logarithmic loss of predictors. *Machine Learning*, 43:247–264, 2001.

Kenny Chiu and Benjamin Bloem-Reddy. Non-parametric hypothesis tests for distributional group symmetry, 2023. arXiv preprint arXiv:2307.15834.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.

Persi Diaconis. Group representations in probability and statistics. *Lecture notes-monograph series*, 11:i–192, 1988.

Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

Morris L. Eaton. Group invariance applications in statistics. *Regional Conference Series in Probability and Statistics*, 1:i–133, 1989. ISSN 19355912. URL http://www.jstor.org/stable/4153172.

Valentina Fedorova, Alex Gammerman, Ilia Nouretdinov, and Volodya Vovk. Plug-in martingales for testing exchangeability on-line. In *Proceedings of the 29th International Conference on Machine Learning*, pages 1639–1646, New York, NY, USA, July 2012. Omnipress. ISBN 978-1-4503-1285-1.

Ronald Aylmer Fisher. Design of experiments. *British Medical Journal*, 1(3923):554, 1936.

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

Ricardo Fraiman, Leonardo Moreno, and Thomas Ransford. Application of the cramér–wold theorem to testing for invariance under group actions. *TEST*, 33(2):379–399, 2024.

Peter D Grünwald and Nishant A Mehta. A tight excess risk bound via a unified pac-bayesian–rademacher–shtarkov–mdl complexity. In *Algorithmic Learning Theory*, pages 433–465. PMLR, 2019.

Peter Grünwald, Rianne de Heide, and Wouter Koolen. Safe Testing. *JRSS B: Statistical Methodology*, 03 2024. ISSN 1369-7412. URL https://doi.org/10.1093/jrsssb/qkae011.

David Haussler and Manfred Opper. Mutual information, metric entropy and cumulative relative entropy risk. *The Annals of Statistics*, 25(6):2451–2492, 1997. ISSN 00905364. URL http://www.jstor.org/stable/2959041.

Alexander Henzi and Michael Law. A rank-based sequential test of independence. *Biometrika*, page asae023, 2024.

John L Kelly. A new interpretation of information rate. *the bell system technical journal*, 35(4):917–926, 1956.

Nick W. Koning. Online permutation tests: e-values and likelihood ratios for testing group invariance, 2023. arXiv preprint arXiv:2310.01153.

Wouter M Koolen and Peter Grünwald. Log-optimal anytime-valid e-values. *Int. Journal of Approximate Reasoning*, 141:69–82, 2022.

Tze Leung Lai. Power-one tests based on sample sums. *The Annals of Statistics*, 5(5): 866–880, 1977.

Eric L Lehmann and Charles Stein. On the theory of some non-parametric hypotheses. *The Annals of Mathematical Statistics*, 20(1):28–45, 1949.

Erich Leo Lehmann, Joseph P Romano, and George Casella. *Testing statistical hypotheses*, volume 3. Springer, 2005.

Shengqiao Li. Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics & Statistics*, 4(1):66–70, 2010.

SV Malov. Sequential ranks and order statistics. *Journal of Mathematical Sciences*, 81: 2434–2441, 1996.

Elizabeth S Meckes. *The random matrix theory of the classical compact groups*, volume 218. Cambridge University Press, 2019.

Muriel Felipe Pérez-Ortiz, Tyron Lardy, Rianne de Heide, and Peter Grünwald. E-statistics, group invariance and anytime valid testing, 2022. arXiv preprint arXiv:2208.07610.

Edwin JG Pitman. Significance tests which may be applied to samples from any populations. *Supplement to the Journal of the Royal Statistical Society*, 4(1):119–130, 1937.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL https://www.R-project.org/.

Aaditya Ramdas, Johannes Ruf, Martin Larsson, and Wouter Koolen. Admissible anytime-valid sequential inference must rely on nonnegative martingales, 2020. arXiv preprint arXiv:2009.03167.

Aaditya Ramdas, Johannes Ruf, Martin Larsson, and Wouter M Koolen. Testing exchangeability: Fork-convexity, supermartingales and e-processes. *International Journal of Approximate Reasoning*, 141:83–109, 2022.

Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 38(4):576–601, 2023.

A Rényi. On the extreme elements of observations. *MTA III. Oszt. Közl*, 12:105–121, 1962.

Herbert Robbins and David Siegmund. The expected sample size of some tests of power one. *The Annals of Statistics*, 2(3):415–436, 1974.

Aytijhya Saha and Aaditya Ramdas. Testing exchangeability by pairwise betting. In *International Conference on Artificial Intelligence and Statistics*, pages 4915–4923. PMLR, 2024.

I. Richard Savage. Contributions to the Theory of Rank Order Statistics-the Two-Sample Case. *The Annals of Mathematical Statistics*, 27(3):590–615, September 1956. ISSN 0003-4851, 2168-8990. doi: 10.1214/aoms/1177728170. Publisher: Institute of Mathematical Statistics.

Pranab Kumar Sen and Malay Ghosh. A chernoff-savage representation of rank order statistics for stationary φ-mixing processes. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 35(2):153–172, 1973a. ISSN 0581-572X. Publisher: Springer.

Pranab Kumar Sen and Malay Ghosh. A Law of Iterated Logarithm for One-Sample Rank Order Statistics and an Application. *The Annals of Statistics*, 1(3):568–576, May 1973b. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1176342426. Publisher: Institute of Mathematical Statistics.

Pranab Kumar Sen and Malay Ghosh. Sequential Rank Tests for Location. *The Annals of Statistics*, 2(3):540–552, May 1974. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1176342714. Publisher: Institute of Mathematical Statistics.

Albert N Shiryaev. *Probability-1*, volume 95. Springer, 2016.

Zbynek Sidak, Pranab K. Sen, and Jaroslav Hajek. *Theory of Rank Tests*. Elsevier, April 1999. ISBN 978-0-08-051910-4.

Adrian FM Smith. On random sequences with centred spherical symmetry. *Journal of the Royal Statistical Society: Series B (Methodological)*, 43(2):208–209, 1981.

Jean Ville. *Etude critique de la notion de collectif*. Gauthier-Villars Paris, 1939.

Vladimir Vovk. On-line confidence machines are well-calibrated. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.*, pages 187–196. IEEE, 2002.

Vladimir Vovk. A universal well-calibrated algorithm for on-line classification. *The Journal of Machine Learning Research*, 5:575–604, 2004.

Vladimir Vovk. The power of forgetting in statistical hypothesis testing. In *Conformal and Probabilistic Prediction with Applications*, pages 347–366. PMLR, 2023.

Vladimir Vovk and Ruodu Wang. E-values: Calibration, combination and applications. *The Annals of Statistics*, 49(3):1736–1754, 2021.

Vladimir Vovk, Ilia Nouretdinov, and Alexander Gammerman. Testing exchangeability on-line. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 768–775, 2003.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*, volume 29. Springer, 2005.

Abraham Wald. *Sequential analysis*. Wiley, New York, 1947.

Appendix A. Proofs

A.1. Proof of Proposition 4

Proof [Proposition 4] Let $F_n(\gamma_{n-1}(X^{n-1}), X_n) = \gamma_n((\gamma_{n-1}(X^{n-1}), X_n))$ for $X^{n-1} \in \mathcal{X}^{n-1}$ and $X_n \in \mathcal{X}$. (With a slight abuse of notation, we refer by $(\gamma_{n-1}(X^{n-1}), X_n)$ to the concatenation of $\gamma_{n-1}(X^{n-1})$ and X_n.) We will show that F_n has the claimed properties. First, we show that $(\gamma_{n-1}(X^{n-1}), X_n)$ and X^n are in the same orbit, so that $\gamma_n((\gamma_{n-1}(X^{n-1}), X_n)) = \gamma_n(X^n)$. To this end, let $g' \in G_{n-1}$ denote the group element such that $g'X^{n-1} = \gamma_{n-1}(X^{n-1})$. Then $\{g(\gamma_{n-1}(X^{n-1}), X_n) : g \in G_n\}$ equals $\{g(g'X^{n-1}, X_n) : g \in G_n\} = \{g\imath_n(g')X^n : g \in G_n\} = \{gX^n : g \in G_n\}$, where we used (iii) of Definition 2 for the first equality and X^n is the concatenation of X^{n-1} and X_n. This shows the first claim.

For the second claim, that $F_n(\,\cdot\,, X_n)$ is one-to-one for each fixed X_n, we show that we can reconstruct $\gamma_{n-1}(X^{n-1})$ from X_n and $\gamma_n(X^n)$. Pick any $g_{X_n} \in G_n$ such that $(g_{X_n}\gamma_n(X^n))_n = X_n$. We know that there exists some $g \in G_n$ such that $gX^n = \gamma_n(X^n)$. Note that $g_{X_n}g$ does nothing to the final coordinate of X^n, so by Item (iii) of Definition 2 there is a $g^*_{n-1} \in G_{n-1}$ such that $g_{X_n}gX^n = \imath(g^*_{n-1})X^n$. We see that $\{\imath(g_{n-1})g_{X^n}\gamma_n(X^n) : g_{n-1} \in G_{n-1}\}$ equals $\{\imath(g_{n-1})g_{X^n}gX^n : g_{n-1} \in G_{n-1}\} = \{\imath(g_{n-1})\imath(g^*_{n-1})X^n : g_{n-1} \in G_{n-1}\} = \{\imath(g_{n-1})X^n : g_{n-1} \in G_{n-1}\}$. It follows from Item (ii) of Definition 2 that $G_{n-1}\mathrm{proj}_{n-1}(g_{X_n}\gamma_n(X^n)) = G_{n-1}X^{n-1}$. It follows that $\gamma_{n-1}(\mathrm{proj}_{n-1}(g_{X_n}\gamma_n(X^n))) = \gamma_{n-1}(X^{n-1})$, as it was to be shown. ■

A.2. Proof of Theorem 8

Proof [Theorem 8] We proceed in two steps: (1) we show that, conditionally on $\gamma_n(X^n)$, R_n is uniformly distributed for each n and (2) we show that $R_1, R_2, \ldots$ are independent. The second step is analogous to the proof of Theorem 3 by Vovk (2002). For each n, define the σ-algebra $\mathcal{G}_n = \sigma(\gamma_n(X^n), X_{n+1}, X_{n+2}, \ldots)$. Notice that $\mathcal{G}_n$ contains, among others, all G_n-invariant functions of X^n because γ_n is a maximally invariant function of X^n—any other G_n-invariant function of X^n is a function of $\gamma_n(X^n)$. Let $g' \in G_n$ such that $\gamma_n(X^n) = g'X^n$, then $\{g \in G_n : A((gX^n)_n, \gamma_n(X^n))_n < \alpha_n\} = \{g \in G_n : A((g\gamma_n(X^n)_n, \gamma_n(X^n))_n < \alpha_n\}g'$–we define $Bg = \{bg : b \in B\}$ for a subset $B \subseteq G_n$. By the invariance of μ_n—it is the Haar probability measure—, it follows that $\mu_n(\{g \in G_n : A((gX^n)_n, \gamma_n(X^n))_n < \alpha_n\}) = \mu_n(\{g \in G_n : A((g\gamma_n(X^n))_n, \gamma_n(X^n))_n < \alpha_n\})$. An analogous identity is derived for the second term in (3). We have $\alpha_n \mid \mathcal{G}_n \overset{\mathcal{D}}{=} A((U\gamma_n(X^n))_n, \gamma_n(X^n))_n \mid \mathcal{G}_n$.

We denote $F(b) := \mu(\{g \in G_n : A((g\gamma_n(X^n))_n, \gamma_n(X^n))_n < b\})$ and define $G(\delta) = \sup\{b \in \mathbb{R} : F(b) \leq \delta\}$. If $\alpha_n \mid \mathcal{G}_n$ is continuous, F is the CDF of that distribution; otherwise, it is the CDF minus the probability of equality. In any case, F is increasing and right-continuous. For any $\delta \in (0,1)$, we have that $F(G(\delta)) = \delta'$ for some $\delta' \leq \delta$, with equality if F is continuous in $G(\delta)$. Then $\mathbb{P}(R_n \leq \delta \mid \mathcal{G}_n) = \mathbb{P}(R_n \leq \delta' \mid \mathcal{G}_n) + \mathbb{P}(\delta' < R_n \leq \delta \mid \mathcal{G}_n)$. Let, for each x, $\Delta F(x) = F(x^+) - F(x)$. For any $\theta \in (0,1]$, we have that $R_n = F(\alpha_n) + \theta\Delta F(\alpha_n) \leq \delta'$ if and only if either $F(\alpha_n) < \delta'$ or $\Delta F(\alpha_n^+) = 0$, which happens precisely when $\alpha_n < G(\delta)$. Therefore, $\mathbb{P}(R_n \leq \delta' \mid \mathcal{G}_n) = \mathbb{P}(\alpha_n < G(\delta') \mid \mathcal{G}_n) = F(G(\delta')) = \delta'$. If F is continuous in $G(\delta)$, then this shows that $\mathbb{P}(R_n \leq \delta \mid \mathcal{G}_n) = \delta$, since $\delta' = \delta$ in that case. If F is not continuous in $G(\delta)$, then $\mathbb{P}(\delta' < R_n \leq \delta \mid \mathcal{G}_n) = \mathbb{P}(\delta' < F(\alpha_n) + \theta\Delta F(\alpha_n) \leq \delta \mid \mathcal{G}_n)$. Notice that $\delta' < F(\alpha_n) + \theta\Delta F(\alpha_n) \leq \delta$ if and only if $\alpha_n = G(\delta)$ and $\theta < (\delta - \delta')/\Delta F(\alpha_n)$, so that

$\mathbb{P}(\delta' < R_n \leq \delta \mid \mathcal{G}_n) = \mathbb{P}(\alpha_n = G(\delta) \mid \mathcal{G}_n)\mathbb{P}\left(\theta \leq \frac{\delta-\delta'}{\Delta F(G(\delta'))} \mid \mathcal{G}_n\right) = (\Delta F(G(\delta')))\frac{\delta-\delta'}{\Delta F(G(\delta'))} = \delta - \delta'$ Putting everything together, we see that $\mathbb{P}(R_n \leq \delta \mid \mathcal{G}_n) = \delta$. This shows the first part, that R_n has a conditional uniform distribution on $[0,1]$.

For the second part of the proof, we show that the sequence $R_1, R_2, \ldots$ is an independent sequence. We have that R_n is $\mathcal{G}_{n-1}$-measurable because it is invariant under transformations of the form $X^n \mapsto (gX^{n-1}, X_n)$ for $g \in G_{n-1}$ (see also Vovk, 2004, Lemma 2). By induction, $\mathbb{P}(R_n \leq \delta_n, \ldots, R_1 \leq \delta_1 \mid \mathcal{G}_n) = \mathbf{E}\left[\mathbf{E}\left[\mathbf{1}\left\{R_n \leq \delta_n, \ldots, R_1 \leq \delta_1\right\} \mid \mathcal{G}_{n-1}\right] \mid \mathcal{G}_n\right] = \mathbf{E}\left[\mathbf{1}\left\{R_n \leq \delta_n\right\}\mathbf{E}\left[\mathbf{1}\left\{p_{n-1} \leq \delta_{n-1}, \ldots, R_1 \leq \delta_1\right\} \mid \mathcal{G}_{n-1}\right] \mid \mathcal{G}_n\right] = \mathbf{E}\left[\mathbf{1}\left\{R_n \leq \delta_n\right\}\right]\delta_{n-1}\cdots\delta_1 = \delta_n \cdots \delta_1$. By the law of total expectation, $\mathbb{P}(R_n \leq \delta_n, \ldots, R_1 \leq \delta_1) = \delta_n \cdots \delta_1$; in other words, $R_1, R_2, \ldots, R_n$ are independent and uniformly distributed on $[0,1]$ for any $n \in \mathbb{N}$. This implies that the distribution of $R_1, R_2, \ldots$ coincides with U^∞ by Kolmogorov's extension theorem (see e.g. Shiryaev, 2016, Theorem II.3.3). ■

A.3. Proof of Proposition 9

The proof of Proposition 9 follows directly from Lemma 11. It states that, with probability one, enough of the original data can be recovered using the smoothed ranks and the orbit representative. We state Lemma 11, prove Proposition 9 and then prove Lemma 11.

Lemma 11 *Suppose, for each $n \in \mathbb{N}$, that $A(\,\cdot\,, \gamma_n(X^n))$ is a one-to-one function of X_n, then there exists a map $D_n : [0,1]^n \times \mathcal{X}^n \to [0,1]^n \times \mathcal{X}^n$ s.t. for any $Q \in \mathcal{H}_0$, $\widetilde{Q}(D_n(R^n, \gamma_n(X^n)) = (\widetilde{\theta}^n, X^n)) = 1$. Here, $\widetilde{\theta}^n = (\widetilde{\theta}_n)_{n\in\mathbb{N}}$ is the sequence given by $\widetilde{\theta}_n = \theta_n \mathbf{1}\left\{\mu_n(\{g \in G_n : A((gX^n)_n, \gamma_n(X^n))_n = \alpha_n\}) \neq 0\right\}$.*

Proof [Proposition 9] Consider, without loss of generality, the case that $A(X^n, \gamma_n(X^n)) = X_n$. By assumption, $R_n \perp \gamma_n$ under P and the marginal distributions of γ_n under Q^* and under P are equal. Ten, $M_n = \frac{\mathrm{d}\widetilde{P}(R^n,\gamma_n(X^n))}{\mathrm{d}\widetilde{Q}^*(R^n,\gamma_n(X^n))}$. Using the sequence of functions $(D_n)_{n\in\mathbb{N}}$ from Lemma 11 and that the external randomization is independent of X^n, the claim follows. ■

Proof [Lemma 11] As in the proof of Theorem 8, we will denote $F(b) = \mu_n(\{g \in G_n : A((gX^n)_n, \gamma_n(X^n))_n < b\})$ and define $G(\delta) = \sup\{b \in \mathbb{R} : F(b) \leq \delta\}$. Furthermore, we will write $\mathbb{P}_{\alpha_n|\gamma_n(X^n)}$ for the distribution of α_n given $\gamma_n(X^n)$ and denote its support by $\mathrm{supp}(\mathbb{P}_{\alpha_n|\gamma_n(X^n)}) := \{x \in \mathbb{R} \mid \text{for all } I \text{ open, if } x \in I \text{ then } \mathbb{P}_{\alpha_n|\gamma_n(X^n)}(I) > 0\}$, If $b \in \mathrm{int}(\mathrm{supp}(\mathbb{P}_{\alpha_n|\gamma_n(X^n)}))$, then there exists an open interval B with $b \in B$ and $B \subseteq \mathrm{supp}(\mathbb{P}_{\alpha_n|\gamma_n(X^n)}))$. For all $c \in B$ with $c > b$, we have that $F(c) - F(b) = \mathbb{P}_{\alpha_n|\gamma_n(X^n)}([b,c)) > 0$, since $[b,c)$ contains an open neighborhood of an interior point of the support. It follows that $F(c) > F(b)$. In words, there are no points c to the right of b such that $F(c) > F(b)$. Consequently, $G(F(b)) = \sup\{a \in \mathbb{R} : F(a) \leq F(b)\} = b$. In a similar fashion, the same identity holds if $b \in \mathrm{supp}(\mathbb{P}_{\alpha_n|\gamma_n(X^n)}) \setminus \mathrm{int}(\mathrm{supp}(\mathbb{P}_{\alpha_n|\gamma_n(X^n)}))$. Notice that $G(R_n) = G(F(\alpha_n) + \theta_n(F(\alpha_n^+) - F(\alpha_n))) = G(F(\alpha_n))$ whenever $\theta_n < 1$, which happens with probability one. Together with the fact that $\mathbb{P}_{\alpha_n|\gamma_n(X^n)}(\mathrm{supp}(\mathbb{P}_{\alpha_n|\gamma_n(X^n)})) = 1$, this gives $\mathbb{P}_{\alpha_n|\gamma_n(X^n)}(G(R_n) = \alpha_n) = 1$, so also $\mathbb{P}(G(R_n) = \alpha_n) = 1$. If $(F(G(R_n)^+) - F(G(R_n))) = \mu_n(\{g \in G_n : (g\alpha^n)_n = \alpha_n\}) = 0$, set $\widetilde{\theta}_n = 0$. If $\mu_n(\{g \in G_n : (g\alpha^n)_n =$

$\alpha_n\}) > 0$, then $\mathbb{P}(\theta_n = (R_n - F(G(R_n)))/(F(G(R_n)^+) - F(G(R_n)))) = 1$, so set $\widetilde{\theta}_n = (R_n - F(G(R_n)))/(F(G(R_n)^+) - F(G(R_n)))$. Since $A(\cdot, \gamma_n(X^n))$ is one-to-one by assumption, its inverse maps α_n to X_n. By Proposition 4, there also exists a map from X_n and $\gamma_n(X^n)$ to $\gamma_{n-1}(X^{n-1})$. At this point, we can repeat the procedure above to recover X_{n-1} from $(R_{n-1}, \gamma_{n-1}(X^{n-1}))$, from which we can then recover $\gamma_{n-2}(X^{n-2})$, etc. Together, all of the maps involved give the function as in the statement of the proposition. ■

A.4. Proof of Theorem 10

Proof [Theorem 10] We first show (6). Assume that $\widetilde{P}$ is such that $R^n \perp \gamma_n(X^n)$ for all n. Let Q^* denote the distribution under which the marginal of $\gamma_n(X^n)$ coincides with that under P, and such that $X^n \mid \gamma_n(X^n) \stackrel{\mathcal{D}}{=} U\gamma_n(X^n) \mid \gamma_n(X^n)$, where $U \sim \mu_n$ and independently of $\gamma_n(X^n)$. First note that $\widetilde{Q}^*\left(M_\tau = \frac{\mathrm{d}P}{\mathrm{d}Q^*}(X^\tau)\right) \geq \widetilde{Q}^*\left(\forall t : M_t = \frac{\mathrm{d}P}{\mathrm{d}Q^*}(X^t)\right) \geq 1 - \sum_{t=1}^{\infty} \widetilde{Q}^*\left(M_t \neq \frac{\mathrm{d}P}{\mathrm{d}Q^*}(X^t)\right) = 1$, where the last inequality uses Lemma 11. By assumption, $\widetilde{P} \ll \widetilde{Q}^*$, and $\widetilde{P}\left(M_\tau = \frac{\mathrm{d}P}{\mathrm{d}Q^*}(X^\tau)\right) = 1$. Consequently, M_τ is a modification of the likelihood ratio evaluated at X^τ.

We show that the likelihood ratio is log-optimal. Let $\ell_n = \frac{\mathrm{d}P}{\mathrm{d}Q^*}(X^n)$ and let $f(\alpha) = \mathbf{E}_{\widetilde{P}}\left[\ln((1-\alpha)\ell_\tau + \alpha E'_\tau)\right]$, a concave function. We show that the derivative of f in 0 is negative, implying that f is maximized at $\alpha = 0$ (our claim). Indeed, $f'(0) = \mathbf{E}_{\widetilde{P}}\left[\frac{E'_\tau - \ell_\tau}{\ell_\tau}\right]$, then $f'(0) = \sum_{i=1}^{\infty} \mathbf{E}_{\widetilde{P}}\left[\frac{E'_i}{\ell_i}\mathbf{1}\{\tau = i\}\right] - 1 = \sum_{i=1}^{\infty} \mathbf{E}_{\widetilde{Q}^*}\left[E'_i \mathbf{1}\{\tau = i\}\right] - 1 = \mathbf{E}_{\widetilde{Q}^*}\left[E'_\tau\right] - 1 \leq 0$, where differentiation and integration were interchanged—the dominated convergence theorem applies because $|f'(\alpha)| = \left|\frac{E'_\tau - \ell_\tau}{(1-\alpha)\ell_\tau + \alpha E'_\tau}\right| \leq \max\left\{\frac{1}{1-\alpha}, \frac{1}{\alpha}\right\}$. Then, $\mathbf{E}_{\widetilde{P}}\left[\ln M_\tau\right] = \mathbf{E}_{\widetilde{P}}\left[\ln E'_\tau\right] \geq \mathbf{E}_{\widetilde{P}}\left[\ln E'_\tau\right]$. Equation (5) follows with the same argument with $\ell'_n = \frac{\mathrm{d}P}{\mathrm{d}Q^*}(R^n)$. ■

Appendix B. Linear Models and Isotropy Groups

Rotational symmetry as described in Section 5.2 describes rotations around the origin. This, we argued, is equivalent to testing whether $X_i \sim \mathcal{N}(0, \sigma)$ for some $\sigma \in \mathbb{R}^+$. In applications it may be more useful to test whether $X^n = \mu\mathbf{1}_n + \epsilon^n$, for $\mu \in \mathbb{R}$ and a spherically symmetric ϵ^n—$\mathbf{1}_n$ is the all-ones n-vector. If μ is known, no additional difficulties are encountered—we can test $X^n - \mu\mathbf{1}_n$ for sphericity. However, the case where μ is unknown can be treated because the null model is still symmetric under a family of rotations. Notice the following: if $O_n \in \mathrm{O}(n)$ is such that $O_n\mathbf{1}_n = \mathbf{1}_n$, then $O_nX^n = \mu\mathbf{1}_n + O_n\epsilon^n$. It follows that $X^n \stackrel{\mathcal{D}}{=} O_nX^n$ every time that $O_n\mathbf{1}_n = \mathbf{1}_n$. That is, the null distribution of X^n is invariant under the isotropy group of $\mathbf{1}_n$; i.e., $G'_n = \{O_n \in \mathrm{O}(n) : O_n\mathbf{1}_n = \mathbf{1}_n\}$. Invariance under the action of G'_n has appeared as centered spherical symmetry (Smith, 1981).

Using geometry, a test is readily obtained. We write $X^n = X^n_{\mathbf{1}_n} + X^n_{\perp\mathbf{1}_n}$, where $X^n_{\mathbf{1}_n} = \frac{\langle X^n, \mathbf{1}\rangle}{n}\mathbf{1}_n$ is the projection of X^n onto the span of $\mathbf{1}_n$; $X^n_{\perp\mathbf{1}_n}$, onto its orthogonal complement. Then, $gX^n = X^n_{\mathbf{1}_n} + gX^n_{\perp\mathbf{1}_n}$ for any $g \in G'_n$. Consequently, the orbit of X^n under G'_n is the intersection of $S^{n-1}(\|X^n\|)$ and the hyperplane $H_n(X^n)$ defined by $H_n(X^n) = \{x'^n \in \mathbb{R}^n : \langle x'^n, \mathbf{1}_n\rangle = \langle X^n, \mathbf{1}_n\rangle\}$. There is a unique line perpendicular to $H_n(X^n)$ that passes through the origin $0_n = (0,\dots,0)$; it intersects $H_n(X^n)$ in the point $0_{H_n} := \frac{1}{n}\langle X^n, \mathbf{1}_n\rangle \mathbf{1}_n$. For any $x'^n \in S^{n-1}(\|X^n\|) \cap H_n(X^n)$, Pythagoras' theorem gives that $\|x'^n - 0_{H_n}\|^2 = \|X^n\|^2 - \|0_{H_n} - 0_n\|^2$. In other words, $S^{n-1}(\|X^n\|) \cap H_n(X^n)$ forms an $(n-2)$-dimensional sphere of radius $(\|X^n\|^2 - \|0_{H_n} - 0_n\|^2)^{1/2}$ around 0_{H_n}. Consider the projection of this sphere on its nth coordinate, normalize the consequent range of values to $[0,1]$, and let $\widetilde{X}_n$ denote the normalized value of X_n. Then R_n is the relative surface area of the $(n-2)$-dimensional hyper-spherical cap with co-latitude angle

$$\varphi = \pi - \cos^{-1}(\widetilde{X}_n / (\|X^n\|^2 - \|0_{H_n} - 0_n\|^2)^{1/2}). \tag{9}$$

Hence, (8) can again be used to determine R_n. A nongeometric derivation using a Gaussian motivation obtained by Vovk (2023), who calls full Gaussian model an online compression model with the summary statistic $\sigma_n = (\langle X^n, \mathbf{1}_n\rangle, \|X^n\|)$.

This model can be extended to cover the presence of covariates; i.e., $X_n = (Y_n, Z_n^d)$ for some $Y_n \in \mathbb{R}$ and $Z_n^d \in \mathbb{R}^d$. Let Z_n be the matrix with row-vectors Z_n^d and assume that Z_n is full rank for every n, a standard assumption in regression. Let $X^n_{\perp Z_n}$ be the projection of X^n onto the complement of the column space of Z_n. The model of interest is $Y^n = Z_n\beta + \epsilon^n$ where $\beta \in \mathbb{R}^d$ and ϵ^n is spherically symmetric. Similar to the reasoning above, this model is invariant under the intersection of the isotropy groups of the column vectors of Z_n; i.e., $G_n = \{O_n \in \mathrm{O}(n) : O_n Z_n = Z_n\}$. The orbit of X^n under G_n is the intersection of $S^{n-1}(\|X^n\|)$ with the column space of Z_n, so that for $\alpha^n(Y^n, Z_n) = Y^n$, computing R_n is analogous. Interestingly, testing for invariance under G_n is not always equivalent to testing for normality with mean $Z_n\beta^d$. A sufficient condition for the equivalence to hold is that $\lim_{n\to\infty}(Z_n' Z_n)^{-1} = 0$; i.e., that β can be estimated consistently via least squares (Eaton, 1989, Section 9.3).

Appendix C. Density estimation in simulations

The density estimation was performed using the kernel density estimation with a Gaussian kernel and as implemented in the `Stats` package in the R language (R Core Team, 2020). Standard kernel density estimation perform poorly around the boundaries—these algorithms are designed to estimate densities supported on $\mathbb{R}$ and not just on $[0,1]$. Following Fedorova et al. (2012), we reflect the sequence $(R_n)_n$ to the left from zero and to the right from one. The density estimate is computed using the extended sample $\cup_{i=1}^n \{-R_i, R_i, 2 - R_i\}$, it is set to zero outside of $[0,1]$, and it is normalized.