Post-hoc and Anytime Valid Inference for Exchangeability and Group Invariance

Nick W. Koning
Econometric Institute, Erasmus University Rotterdam, the Netherlands

March 5, 2024

Abstract

We study post-hoc (e-value-based) and post-hoc anytime valid inference for testing exchangeability and general group invariance. Our methods satisfy a generalized Type I error control that permits a data-dependent selection of both the number of observations n and the significance level α. We derive a simple analytical expression for all exact post-hoc valid p-values for group invariance, which allows for a flexible plug-in of the test statistic. For post-hoc anytime validity, we derive sequential p-processes by multiplying post-hoc p-values. In sequential testing, it is key to specify how the number of observations may depend on the data. We propose two approaches, and show how they nest existing efforts. To construct good post-hoc p-values, we develop the theory of likelihood ratios for group invariance, and generalize existing optimality results. These likelihood ratios turn out to exist in different flavors depending on which space we specify our alternative. We illustrate our methods by testing against a Gaussian location shift, which yields an improved optimality result for the t-test when testing sphericity, connections to the softmax function when testing exchangeability, and an improved method for testing sign-symmetry.

Keywords: permutation test, group invariance test, anytime valid inference, post-hoc valid inference, e-values, sequential testing.

Contents

1 Introduction **3**
- 1.1 Methodological contributions . . . 3
- 1.2 Technical Contributions . . . 4
- 1.3 Related literature . . . 6
- 1.4 Notation and underlying assumptions . . . 7

2 Valid and post-hoc valid group invariance tests **7**
- 2.1 Group invariance . . . 7
- 2.2 Orbit Independence . . . 8
- 2.3 Traditional test for group invariance . . . 9
- 2.4 Post-hoc group invariance tests . . . 10
- 2.5 Obtaining the normalization constant . . . 11

3 Post-hoc anytime valid group invariance testing **11**
- 3.1 Sequential invariance . . . 11
- 3.2 Post-hoc anytime valid p-process for group invariance . . . 13

4 Modifying the filtration **14**
- 4.1 First approach: invariance of a statistic . . . 15
- 4.2 Second approach: invariance through a statistic . . . 16

5 Likelihood ratios for group invariance **18**
- 5.1 Alternative on the sample space $\mathcal{Y}$ . . . 19
- 5.2 Alternative on the orbits . . . 20
- 5.3 Alternative on the group $\mathcal{G}$ and inversion kernels . . . 21

6 Illustration: LRs for invariance vs Gaussianity **21**
- 6.1 Sphericity . . . 22
- 6.2 Likelihood ratio on $\mathcal{Y}$ . . . 22
- 6.3 Likelihood ratio on orbits . . . 23
- 6.4 Likelihood ratio on $\mathcal{G}$ . . . 24
- 6.5 Permutations and softmax . . . 24
- 6.6 Testing sign-symmetry . . . 25

7 Simulations **25**
- 7.1 Case-control experiment and learning the alternative . . . 25
- 7.2 Testing symmetry and a comparison to de la Peña [1999] . . . 26

8 Acknowledgements **28**

A Conditional distributions **29**

B Example: LR for exchangeability **30**
- B.1 Exchangeability on a finite sample space . . . 30
- B.2 Likelihood ratios . . . 31

C Finding the normalization constant analytically **32**

D Omitted proofs **32**
- D.1 Proof of Theorem 1 . . . 32
- D.2 Proof of Theorem 2 . . . 33
- D.3 Proof of Theorem 3 . . . 34
- D.4 Proof of Theorem 4 . . . 35
- D.5 Proof of Proposition 2 . . . 35
- D.6 Proof of Theorem 5 . . . 35

D.7 Proof of Theorem 6 36
D.8 Proof of Lemma 1 36
D.9 Proof of Proposition 3 37

1 Introduction

Suppose we observe $X^n = X_1, \ldots, X_n$ and are interested in testing whether these objects are exchangeable:

$$H_0 : X^n \text{ is exchangeable.}$$

Exchangeability here means that X^n is equal in distribution to any permutation PX^n of its elements. For example, X^n is exchangeable if $X_1, \ldots, X_n$ are i.i.d.

Traditionally, we want to test this hypothesis at some pre-specified level α. To construct a so-called permutation test, we select a test statistic T and reject whenever the following permutation p-value is at most α:

$$p(X^n) = \mathbb{P}_{\overline{P}_n}\left(T(\overline{P}_n X^n) > T(X^n)\right) + \overline{u}\mathbb{P}_{\overline{P}_n}\left(T(\overline{P}_n X^n) = T(X^n)\right),$$

where $\overline{P}_n \sim \text{Unif}(\mathfrak{P}_n)$ is uniformly distributed on the permutations $\mathfrak{P}_n$ and $\overline{u}$ is independently uniform on $[0, 1]$. This p-value can be understood as the proportion of test statistics calculated from the rearranged ('permuted') data that exceed or match the original test statistic, and a small correction to handle discreteness. The resulting test is well-known to be exactly valid under the Type I error in finite samples:

$$\sup_{n,\alpha} \mathbb{E}_{X^n}\left[\mathbb{I}\left\{p(X^n) \leq \alpha\right\}/\alpha\right] = 1, \tag{1}$$

or, equivalently, $\mathbb{P}_{X^n}(p(X^n) \leq \alpha) = \alpha$, for all n and α. Permutation tests are a special case of more general group invariance tests, which are obtained by simply replacing the group of permutations $\mathfrak{P}_n$ by another compact group $\mathcal{G}_n$ that acts on our sample space.

1.1 Methodological contributions

The first unfortunate feature of this traditional group invariance test is that the significance level α must be pre-specified. Our first methodological contribution is to introduce a 'post-hoc' p-value $\mathfrak{p}$ for group invariance that is valid under a much stronger Type I error property [Grünwald, 2023, Koning, 2024], also called the post-hoc Type I error:

$$\sup_{n} \mathbb{E}_{X^n}\left[\sup_{\alpha} \mathbb{I}\{\mathfrak{p}(X^n) \leq \alpha\}/\alpha\right] \equiv \mathbb{E}_{X^n}\left[1/\mathfrak{p}(X^n)\right] = 1. \tag{2}$$

This is stronger than the Type I error in (1), because the supremum over α is now inside the expectation so that α can be selected based on the data. The reciprocal of such a 'post-hoc valid' p-value also known as an e-value [Shafer, 2021, Vovk and Wang, 2021, Howard et al., 2021, Ramdas et al., 2023, Grünwald et al., 2023].

In particular, we show that every exactly post-hoc valid p-value for group invariance is of the form

$$\mathfrak{p}(X^n) = \frac{\mathbb{E}_{\overline{G}} T(\overline{G} X^n)}{T(X^n)},$$

for some non-negative and appropriately integrable test statistic T, where $\overline{G} \sim \text{Unif}(\mathcal{G}_n)$.

A second unfortunate feature is that the number of observations n must also be pre-specified, or at least specified independently from the data. Our second methodological contribution is to solve this, by introducing post-hoc anytime valid p-processes $(\mathfrak{p}_n)_{n\geq 1}$ for testing group invariance. Such p-processes are a sequential generalization of post-hoc p-values, which bound the post-hoc Type I error even if n is a potentially data-dependent stopping time:

$$\sup_{n \in \mathfrak{N}} \mathbb{E}_{X^n}\left[\sup_{\alpha} \mathbb{I}\{\mathfrak{p}_n(X^n) \leq \alpha\}/\alpha\right] \equiv \sup_{n \in \mathfrak{N}} \mathbb{E}_{X^n}\left[1/\mathfrak{p}_n(X^n)\right] \leq 1,$$

where $\mathfrak{N}$ is a collection of stopping times adapted to the available information at each moment in time. This is a stronger property than (2), as it allows us to dynamically decide when we are done with collecting data based on our current information, instead of having to independently specify a number of observations. The reciprocal of a p-process that satisfies this property also known an e-process [Ramdas et al., 2022a, 2023].

In particular, we consider invariance of a sequence of data $(X^n)_{n\geq 1}$ under a sequence of groups $(\mathcal{G}_n)_{n\geq 1}$. We then introduce post-hoc anytime valid p-processes of the form

$$\mathfrak{p}(X^n) = \prod_{i=1}^{n} \frac{\mathbb{E}_{\overline{F}_i} T_i(\overline{F}_i X^i)}{T_i(X^i)},$$

where $\overline{F}_i \sim \text{Unif}(\mathcal{F}_i)$, and $\mathcal{F}_i$ is a compact subgroup of $\mathcal{G}_i$ that is adapted to the available information at every moment, $i \leq n$. Here, the test statistics $(T_n)_{n\geq 1}$ can be also chosen based on the available information. These p-processes can be interpreted as the product of post-hoc p-values. Their reciprocal also called an e-process, and is a martingale.

Finally, we formulate how the test statistic T can be chosen optimally. For an alternative $\mathbb{P}_1$, a popular post-hoc generalization of power is $\mathbb{E}^{\mathbb{P}_1} \log 1/\mathfrak{p}$ [Koolen and Grünwald, 2022, Grünwald et al., 2023]. We argue that under this notion of power, the optimal choice is to choose T equal to the density under the alternative $\mathbb{P}_1$. The resulting post-hoc p-value can be interpreted as a likelihood ratio between $\mathbb{P}_1$ and $\mathcal{G}$ invariance. In the sequential setting, this alternative can be learned based on the previously observed data.

1.2 Technical Contributions

Beyond the methodological contributions, we make several technical contributions to testing group invariance. To start, we introduce a novel weaker condition under which inference on group invariance can take place. Rather than the data being invariant, we specify a condition that merely requires it to be invariant when viewed through a test statistic. To the best of our knowledge, this generalizes the weakest known condition that appears in Hemerik and Goeman [2018]. Based on this condition, we develop the theory of for Type I and post-hoc Type I valid inference for group invariance. Here, we derive the class of all exactly post-hoc valid p-values for group invariance.

Next, we setup a general framework for sequential testing of group invariance. We then first develop post-hoc anytime valid p-processes based on martingales for the 'standard' filtration. This standard filtration allows us to fully inspect all the available data up until the present to determine when we want to stop gathering new observations. Moreover, our methods permit us flexibly choose the test statistic based on the available data, for example to learn the alternative as the data comes in.

For the special case of testing sequential exchangeability, Ramdas et al. [2022a] and Vovk [2021] observe that relying on the standard filtration can be problematic when using a martingale: the resulting procedure becomes powerless. We give a deep explanation of why this problem occurs when testing group invariance more generally. To solve this issue, we consider reducing the filtration. This amounts to only allowing the selection of n and the test statistic to depend on a statistic of the previous data. We propose two approaches. The first approach is simplest, and relies on reducing the problem to testing invariance of a statistic of the data. The second approach is more general, and does not assume we have a group acting on the codomain of the statistic. Instead, we rely on our previously mentioned condition that only requires the data to be invariant when viewed through the statistic. We show how this second approach nests the approaches of Vovk [2021] and its recent generalizations by Lardy and Pérez-Ortiz [2024]. Given the complexity of this problem, we view this as our primary technical contribution.

Furthermore, we consider how to construct optimal post-hoc p-values. There are several competing generalizations of 'power' in post-hoc inference. For example, for an alternative $\mathbb{P}_1$, we may either want to maximize $\mathbb{E}^{\mathbb{P}_1} \log 1/\mathfrak{p}$ or maximize $\mathbb{E}^{\mathbb{P}_1} 1/\mathfrak{p}$ [Koolen and Grünwald, 2022, Grünwald et al., 2023, Koning, 2024]. In traditional inference, likelihood ratio tests are known to yield optimal power by the Neyman-Pearson lemma. Interestingly, likelihood ratios also play a central role in optimal post-hoc inference under such generalizations of power. For example, when maximizing $\mathbb{E}^{\mathbb{P}_1} \log 1/\mathfrak{p}$, the optimal choice is the likelihood ratio $1/\mathfrak{p} = d\mathbb{P}_1/d\mathbb{P}_0$, for a simple null $\mathbb{P}_0$. This choice has the additional benefit that their reciprocal can simultaneously be interpreted as a post-hoc p-value under the alternative [Koning, 2024].

For these reasons, we develop likelihood ratio statistics for group invariance. Surprisingly, this is possible, even though group invariance is a highly composite null hypothesis. To accomplish this, we reduce the hypothesis to simple null hypotheses on certain non-overlapping maximally invariant subsets (orbits) of the sample space, and treat these separately. We show that such likelihood ratio statistics are easy to construct, by simply choosing T equal to the density under the alternative.

We find these likelihood ratios come in three distinct flavors, that depend on the space on which the alternative is specified. The first flavor relies on specifying an alternative on the sample space, and is inspired by an old proof strategy of Lehmann and Stein [1949] for Neyman-Pearson optimality in traditional inference. For the second flavor, we instead specify an alternative on the invariant subsets (orbits). As a side-contribution, we explain how we can use this second flavor to generalize the optimality result of Lehmann and Stein [1949] to certain composite alternatives. The third flavor relies on specifying an alternative on the group itself. For this flavor, we use an inversion kernel (see Ch. 7 in Kallenberg 2017), which was recently introduced in the context of group invariance testing by Chiu and Bloem-Reddy [2023]. A toy example of these different types of likelihood ratios can be found in Appendix B.

We illustrate these likelihood ratio statistics to testing invariance on $\mathbb{R}^d$, $d \geq 1$, under an arbitrary group of orthonormal matrices, against a simple alternative that is a location shift under normality. For the special case of spherical invariance, this is connected to an example from Lehmann and Stein [1949] regarding the optimality of the t-test, which we slightly generalize. We also consider sign-symmetry, which produces a post-hoc p-value that can be viewed as an admissible version of a post-hoc p-value based on de la Peña [1999]. Furthermore, we consider exchangeability where we find that the softmax function is nested as a special case of our likelihood ratio statistic.

We empirically illustrate our methods in two simulation experiments. The first experiment mimics a standard case-control experiment under random treatment allocation. In the second experiment we compare our sign-symmetry p-process to the one based on de la Peña [1999], and find that it is dramatically more powerful.

1.3 Related literature

At first glance, our work may seem intimately related to the work of Pérez-Ortiz et al. [2022]. However, they consider invariance of *collections of distributions* (both the null and the alternative), whereas we consider invariance of *distributions themselves*. Specifically, a collection of distributions $\mathcal{P}$ is said to be invariant under a transformation g if for any $\mathbb{P} \in \mathcal{P}$, its transformation $g\mathbb{P}$ by g is also in $\mathcal{P}$. In contrast, invariance of a distribution $\mathbb{P}$ means that its transformation $g\mathbb{P}$ is equal to $\mathbb{P}$ itself. Intuitively, their work can be interpreted as *testing in the presence of an invariant model*, whereas we consider *testing the invariance of the data generating process*.

As our null hypothesis consists exclusively of invariant distributions it is technically also invariant, so that one may believe their results may still apply under appropriate assumptions on the alternative. However, this invariance is of a very strong type which excludes the transitivity that Pérez-Ortiz et al. [2022] require. In some sense, the strong type of invariance we consider is the complete opposite of transitivity.

Vovk [2023b] independently derives a permutation test for e-values, or equivalently, post-hoc valid p-values. However, his work only considers a toy example which tests exchangeability of binary data for a single specific alternative hypothesis. Our work generalizes this to invariance under a compact group of locally compact Hausdorff valued data against arbitrary alternatives and using arbitrary test statistics.

A closely related work is that of Ramdas et al. [2022b], who consider testing sequential exchangeability. However, they focus primarily on the case where $X = X_1, X_2, \dots$ is a binary or d-ary sequence. Their methods rely on taking an infimum over multiple non-negative supermartingales, which itself is no longer a supermartingale, but is still an e-process. It is not clear how their methods can be extended beyond d-ary data, nor to other forms of group invariance. Moreover, Vovk [2023a] argues that this approach does not generalize beyond "toy situations".

There is more related work by Vovk [2021], which considers sequential exchangeability. He exploits the fact that the sequential ranks are independent from the past ranks under sequential exchangeability. He then converts these ranks into independent post-hoc p-values, which are multiplied together to construct a post-hoc p-process. Lardy and Pérez-Ortiz [2024] generalize the approach of Vovk [2021] to sequential group invariance. In Section 4.2 we propose a more general framework and explain how their approaches are nested in ours. The key improvement is that our approach does not require the use of ranks, and more abstractly does not require the test statistic to have the same distribution on every orbit.

A link between the softmax function and post-hoc p-values for exchangeability was also made in unpublished early manuscripts of Wang and Ramdas [2022] and Ignatiadis et al. [2023], which they call a 'soft-rank' e-value. In Remark 4, we explore the connection to our softmax likelihood ratio statistic, and find that their soft-rank e-value can be interpreted as a more variable version.

1.4 Notation and underlying assumptions

Throughout the paper, whenever we define a sample space we consider it to be second countable locally compact Hausdorff, equipped with a Borel σ-algebra. Moreover, the groups we consider are second countable compact Hausdorff topological groups, which we will just refer to as 'compact groups' that act continuously on the associated sample space.

To avoid ambiguity, we sometimes write expectations $\mathbb{E}$ with a superscript and/or subscript $\mathbb{E}_X^{\mathbb{P}}$ to make explicit the measure over which is being integrated ($\mathbb{P}$), and the random variables over which the integration takes place (X). We use similar subscripts for probabilities.

2 Valid and post-hoc valid group invariance tests

2.1 Group invariance

Let $\mathcal{G}$ be a compact group, acting continuously on some sample space $\mathcal{Y}$. Examples of compact groups acting on $\mathbb{R}^n$ include rotations, and permutations and sign-flips that act on the canonical basis vectors. Such groups can be represented by collections of orthonormal matrices that are closed under matrix multiplication and inverses, and act on $\mathbb{R}^n$ through matrix multiplication.

The 'orbit' of $y \in \mathcal{Y}$, denoted by $O_y = \{z \in \mathcal{Y} \mid z = Gy,\ \exists G \in \mathcal{G}\}$, can be interpreted as the set of all points that can be reached when starting from y and applying an element of the group to it. We assign a single point $[y]$ on each orbit as the 'orbit representative' of O_y. That is, $[y] = Gy$ for some $G \in \mathcal{G}$. We use $[\mathcal{Y}]$ to denote the collection of orbit representatives, and $\mathcal{Y}/\mathcal{G}$ for the collection of all orbits, and we call the function $[\cdot] : \mathcal{Y} \to [\mathcal{Y}]$ that maps y to its orbit representative $[y]$ an orbit selector.

We say that a random variable Y on $\mathcal{Y}$ is $\mathcal{G}$ invariant if its law remains unchanged after a transformation by any element of $\mathcal{G}$.

Definition 1 (Invariance)**.** *A random variable Y is $\mathcal{G}$ invariant if $Y \overset{d}{=} GY$, for all $G \in \mathcal{G}$.*

Alternatively, we can say that the conditional distribution of Y given $Y \in O$ is uniform on O, for (almost) every orbit O. That is, the law of a $\mathcal{G}$ invariant random variable is some mixture of uniform distributions on orbits.

Equivalently, we can say $Y \overset{d}{=} \overline{G}Y$, where $\overline{G}$ is uniform (Haar) distributed on $\mathcal{G}$ independently from Y. Moreover, it is equivalent to $Y \overset{d}{=} \overline{G}[Y]$, provided that the orbit selector is measurable. This means that an invariant random variable can be decomposed or *deconvolved* into a uniform random variable on the group multiplied (using the group action) by a random variable that is distributed over orbit representatives (see e.g. Eaton 1989).

Sometimes, we only look at the data through a statistic S, that maps to some space $\mathcal{H}$. Using such a statistic, we can define the following weaker notion of invariance. The standard notion of invariance is recovered if S is invertible.

Definition 2 (Invariance through a statistic)**.** *A random variable Y is $\mathcal{G}$ invariant through S if conditional on $Y \in O$, $S(\overline{G}Y) \stackrel{d}{=} S(Y)$, where $\overline{G}$ is uniform on $\mathcal{G}$ independently from Y, for all $O \in \mathcal{Y}/\mathcal{G}$.*

We illustrate the difference between invariance and conditional invariance in two examples. In Example 1, the random variable is not exchangeable, but exchangeable through a statistic. In Example 2, the random variable is exchangeable.

Example 1 (Invariant through a statistic)**.** *Suppose we have two bags. We fill one with the numbers 1 and 2, and the other with numbers 3 and 4. We now first pick a bag with equal probability, and then sequentially draw both numbers from the bag in an exchangeable manner. Next, we take the other bag and do the same, after which we arrange the numbers in the order they were drawn.*

While the choice of bag is exchangeable, and the order of the numbers drawn from each bag is exchangeable, the resulting set of numbers is not exchangeable. For example, the order 1423 cannot occur, since 1 and 4 are not in the same bag and can therefore never be drawn as the first two numbers. For this same reason, it is also not exchangeable through the statistic that returns the first two elements. On the other hand, the order is exchangeable through the statistic that returns the value of the first drawn number, as the probability for any number to be drawn first is equal.

Abstractly speaking, our group here is the permutation subgroup that permutes within the bags and permutes the two bags themselves: the permutations that permute numbers between bags are not in this subgroup. There is a single orbit: $\{1234, 2134, 1243, 2143, 3412, 3421, 4312, 4321\}$, *and our process draws uniformly from this orbit.*

Example 2 (Unconditionally invariant through a statistic)**.** *We now consider a different thought experiment, where we stop after the first bag. Suppose we again have two bags, and again fill one with the numbers 1 and 2, and the other with numbers 3 and 4. We first choose a single bag with an arbitrary probability distribution, and put the other bag away. Then, we sequentially draw the two numbers exchangeably from this bag and arrange them in the order they were drawn.*

Suppose we consider a statistic S that converts the numbers into their ranks: $S(12) = 12$, $S(21) = 21$, $S(34) = 12$, $S(43) = 21$. The resulting ranks are certainly exchangeable. More interestingly, the distribution of the ranks is the same for every bag. As a consequence, the ranks are not just uniform conditionally on every orbit, but also uniform unconditionally.

Abstractly speaking, our sample space is $\{12, 21, 34, 43\}$ and our group is the permutation group on 2 units. The orbits are $\{12, 21\}$ and $\{34, 43\}$, so that the selection of the bag can be viewed as the selection of the orbit.

2.2 Orbit Independence

In Example 2, the statistic is chosen in such a way that the random variable is identically distributed on every orbit. That is, the distribution of $T(Y)$ is independent from the orbit.

More examples of statistics on $\mathbb{R}^n$ that have the same distribution for every orbit are ranks for exchangeability, signs for sign-symmetry, and a normalized vector $Y/\|Y\|_2$ for spherical invariance (assuming no ties in the ranks, zeroes for the signs, and excluding the zero-vector for the spherical invariance). Functions of these statistics give rise to rank tests, sign tests and the t-test (see Example 3). These particular statistics are also maximal choices in some sense: the group actions commute with these statistics. For example, $GY/\|GY\|_2 = G(Y/\|Y\|_2)$ for any orthonormal matrix G.

A general method to construct statistics with the same distribution on every orbit is by using an inversion kernel $\gamma : \mathcal{Y} \to \mathcal{G}$ [Kallenberg, 2017], which maps data to an element of the group. This effectively makes our sample space equal to the group $\mathcal{Y} = \mathcal{G}$, and since a group acts transitively on itself there is just one orbit to consider. In turn, this means that the distribution of $\gamma(Y)$ is necessarily the same on every orbit, since there is only one. More detail on such inversion kernels is given in Section 5.3. The previous examples are closely related to inversion kernels, as the ranks (barring ties) and signs (barring zeroes) are in bijective correspondence with the groups of permutations and sign-flips. Moreover, the normalized vector $Y/\|Y\|_2$ can be formulated as a (non-bijective) function $\gamma(Y)\iota$ of an inversion kernel for the orthogonal group, where ι is an arbitrary unit vector.

2.3 Traditional test for group invariance

In this section, we consider the traditional Type I error control. It is straightforward to generalize the permutation test in Section 1 to a so-called group invariance test that also controls the Type I error exactly. Such a group invariance test tests whether a random variable is invariant under a compact group, which specializes to exchangeability for the permutation group.

Formulated in terms of orbits, we wish to test whether the conditional distribution of the random variable is uniform on every orbit. The way a group invariance test works is to test whether the random variable is uniform on the orbit in which it landed; if we can reject this, we can reject that it is uniform on every orbit. Moreover, as we only observe data on one orbit, this is also all we can do.

Given any test statistic $T : \mathcal{Y} \to \mathbb{R}$, designed to be large under the alternative, an exactly valid p-value can be defined as

$$p(Y) = \mathbb{P}_{\overline{G}}\left(T(\overline{G}Y) > T(Y)\right) + \overline{u}\mathbb{P}_{\overline{G}}\left(T(\overline{G}Y) = T(Y)\right), \tag{3}$$

where $\overline{G} \sim \text{Unif}(\mathcal{G})$ is uniformly (Haar) distributed on $\mathcal{G}$, which is well-defined as $\mathcal{G}$ is compact, and $\overline{u}$ is uniform on $[0,1]$. Comparing this p-value to a pre-specified significance level $\alpha \in (0,1]$ yields a Type I valid test. This test is equivalent to rejecting $H_0^{\mathcal{G}}$ if

$$T(Y) > q_\alpha^{\overline{G}}[T(\overline{G}Y)], \tag{4}$$

and with some appropriate probability in case of equality, where $q_\alpha^{\overline{G}}[T(\overline{G}Y)]$ denotes the α upper-quantile of the distribution of $T(\overline{G}Y)$ where Y is considered fixed.

If Y is $\mathcal{G}$ invariant, it is well-known that this test and its associated p-value are valid for any test statistic T. In fact, we show this also holds if Y is $\mathcal{G}$ invariant through T, as in Definition 2. To the best of our knowledge, this aspect is novel. The t-test, which is an example of a group invariance test, is given in Example 3. Example 4 covers the most basic form of conformal inference [Shafer and Vovk, 2008].

Theorem 1. *If Y is $\mathcal{G}$ invariant through T, then $p(Y)$ is uniform on* $[0,1]$.

Example 3 (t-test)**.** *Suppose $\mathcal{Y} = \mathbb{R}^n$ and T is defined as $T(y) = \iota' y/\|y\|_2$, where ι is some unit vector. If Y is spherically invariant through T, then $T(Y)$ is* Beta$(\frac{n-1}{2}, \frac{n-1}{2})$*-distributed on* $[-1, 1]$ *(see e.g. Koning and Hemerik [2023] for a proof) conditional on every orbit, and so unconditionally as well. Equivalently, $\sqrt{n-1}T(Y)/\sqrt{1-T(Y)^2}$ is t-distributed. The resulting test for spherical invariance is also known as the t-test.*

Example 4 (Conformal inference)**.** *Suppose $\mathcal{Y} = \mathbb{R}^{n+1}$ and $\mathcal{G}$ is the group of permutations acting on the canonical basis of $\mathbb{R}^{n+1}$. Let Y^{n+1} be an exchangeable random variable, and let $T : \mathcal{Y} \to \mathbb{R}$ be a test statistic that only depends on the final element Y_{n+1}. Suppose we only observe Y^n and want to test whether the unobserved Y_{n+1} could be equal to some hypothesized value y^*. We can then use the permutation test based on $T((Y^n, y^*))$, which is also known as conformal inference. Repeating this test for all $y^* \in \mathcal{Y}$ and collecting the values of y^* for which we do not reject yields the conformal prediction set, which is a confidence set for Y_{n+1} on $\mathbb{R}$.*

2.4 Post-hoc group invariance tests

In this section, we derive tests for group invariance that are not just valid, but post-hoc valid, as defined in (2). As with the traditional Type I valid test for group invariance treated in Section 2.3, we still have great freedom in our selection of the test statistic for post-hoc testing. In particular, let $T : \mathcal{Y} \to \mathbb{R}_+$ be some arbitrary non-negative test statistic that is appropriately integrable on every orbit $O \in \mathcal{Y}/\mathcal{G}$. Namely $0 < \mathbb{E}_{\overline{G}} T(\overline{G}y) < \infty$ for every $y \in \mathcal{Y}$. Based on this test statistic, we consider as p-value

$$\mathfrak{p}_T(Y) = \frac{\mathbb{E}_{\overline{G}} T(\overline{G}Y)}{T(Y)}, \tag{5}$$

where $\overline{G} \sim \text{Unif}(\mathcal{G})$. The interpretation is that $\mathfrak{p}_T(Y)$ is small if $T(Y)$ is large compared to its average value on the orbit of Y. Moreover, as we shall show in Section 5, $\mathfrak{p}_T$ can be interpreted as a likelihood ratio for $\mathcal{G}$ invariance against a density proportional to T.

In Theorem 2 we not only show that these p-values are exactly post-hoc valid, but also the converse: any exactly post-hoc p-value can be written as in (5). The result is not obvious, as $\mathcal{G}$ invariance is a large composite hypothesis, and the numerator in (5) only takes the expectation over the group. Its proof can be found in Appendix D.2.

Theorem 2. *$\mathfrak{p}_T(Y)$ is exactly post-hoc valid if Y is $\mathcal{G}$ invariant through T. Conversely, any exact post-hoc p-value for $\mathcal{G}$ invariance through a statistic is of this form for some T.*

By Theorem 2, we can use any appropriately integrable test statistic T to construct an exact post-hoc p-value for $\mathcal{G}$ invariance. In fact, as a non-exact post-hoc p-values is a statistic, we can plug it in for T to transform it into an exact variant. We exploit this trick in Section 6.6. Moreover, as this class contains all exact post-hoc p-values, we are not 'missing' any important post-hoc p-values.

Example 5 (Is there a post-hoc t-test?)**.** *Continuing from Example 3, it is unfortunately not clear how to generalize the t-test to a post-hoc test: there are many possible candidates. For example, we could derive one using the statistic $T(y) = \exp\{\iota' y/\|y\|_2\}$, but any non-negative strictly increasing function of $\iota' y/\|y\|_2$ could reasonably qualify. The key underlying* *problem is that the normalization in* (5) *is only invariant to scalar-transformations of the statistic, and not invariant to strictly increasing transformations of the statistic like the quantile function in traditional tests such as* (4)*.*

In Section 6, we offer an alternative generalization of the t-test, by re-interpreting the t-test as a likelihood ratio test for spherical invariance against a Gaussian alternative. Another potential generalization is offered by Pérez-Ortiz et al. [2022], who start with a Gaussian model and integrate out the variance.

Example 6 (Post-hoc conformal inference)**.** *Continuing the set-up from Example 4, if T is a non-negative test statistic that only depends on the final element, then $\mathbb{E}_{\overline{P}}T(\overline{P}(Y^n, y^*))/T((Y^n, y^*))$ is a post-hoc p-value for conformal inference.*

2.5 Obtaining the normalization constant

The main computational challenge when using post-hoc p-values for group invariance is the computation of the normalization constant $\mathbb{E}_{\overline{G}}T(\overline{G}Y)$. As the group $\mathcal{G}$ is often large, simply averaging $T(GY)$ over all G may not be feasible. However, the normalization constant can be estimated.

We borrow some ideas from traditional group invariance tests, where similar issues occur. The simplest idea is to use a Monte Carlo approach by replacing $\overline{G}$ with a random variable $\overline{G}^M$ that is uniformly distributed on a set of i.i.d. draws $\{\overline{G}^{(1)}, \overline{G}^{(2)}, \dots, \overline{G}^{(M)}\}$ of $\overline{G}$. Alternatively, we can replace $\overline{G}$ with $\overline{H}$ that is uniformly distributed on a compact subgroup of $\mathcal{G}$ [Chung and Fraser, 1958]. As invariance under $\mathcal{G}$ implies invariance under every subgroup, this still guarantees the resulting p-value is post-hoc valid even if the subgroup is small, which is not clear for the Monte Carlo approach if the number of samples is small. Such a subgroup may also be easier to work with than $\mathcal{G}$ itself. Moreover, Koning and Hemerik [2023] note that we can actually strategically select the subgroup based on the test statistic and alternative, and select a subgroup that yields high power. Koning [2023] observes that this can even yield testing methods that are more powerful than if we use the entire group $\mathcal{G}$.

Note that in the traditional group invariance test, the goal is to estimate the α-upper quantile the distribution of $T(Y)$ given $Y \in O_Y$, as in (4). The normalization constant is the mean of this same distribution, which we expect to be much easier to estimate in practice. Based on simulation results, it seems that roughly 100 draws is usually sufficient. Moreover, in Appendix C we discuss that we can sometimes very efficiently approximate the normalization constant analytically. In addition, if $T(Y)$ is orbit independent if Y is $\mathcal{Y}$ invariant, then the distribution can even be pre-computed, at is it not necessary to know the orbit O_Y of Y.

3 Post-hoc anytime valid group invariance testing

3.1 Sequential invariance

We start with describing the sample space. We embed the entire sequential setting in a latent sample space $\mathcal{X}$. In particular, we assume we have a nested sequence of subsets

$(\mathcal{X}^n)_{n\geq 1}$ of $\mathcal{X}$: $\mathcal{X}^n \subseteq \mathcal{X}^{n+1}$, which are tied together through a sequence of continuous projection maps $(\text{proj}_{\mathcal{X}^n})_{n\geq 1}$, $\text{proj}_{\mathcal{X}^n} : \mathcal{X} \to \mathcal{X}^n$.

To describe the sequence of data we observe, we assume there is some latent random variable X on $\mathcal{X}$, of which we sequentially observe an increasingly rich sequence $(X^n)_{n\geq 1}$ of projections $X^n = \text{proj}_{\mathcal{X}^n}(X)$, $n \geq 1$.[1] This construction ensures that this sequence of random variables induces a filtration $(\sigma(X^n))_{n\geq 1}$.

Next, we consider the group structure. Our sequential group structure is embedded into a compact group $\mathcal{G}$ that acts continuously on $\mathcal{X}$. In particular, we consider a sequence of subgroups $(\mathcal{G}_n)_{n\geq 1}$ of $\mathcal{G}$. We assume the projection map induces a group action of $\mathcal{G}_n$ on $\mathcal{X}^n$ through the group action on $\mathcal{X}$: $GX^n = \text{proj}_{\mathcal{X}^n}(GX)$, for all $G \in \mathcal{G}_n$.[2] This assumptions ensures we can use the groups $(\mathcal{G}_n)_{n\geq 1}$ and observations $(X^n)_{n\geq 1}$ without reference to the latent $\mathcal{G}$, $\mathcal{X}$ and X.

We are now ready to define our notion of sequential invariance.

Definition 3 (Sequential invariance)**.** *$(X^n)_{n\geq 1}$ is $(\mathcal{G}_n)_{n\geq 1}$ invariant if X^n is $\mathcal{G}_n$ invariant for all n.*

Equivalently, we can define sequential invariance as the conditional distribution of X^n being uniform on every orbit $O \in \mathcal{X}^n/\mathcal{G}_n$, for each n. As X^n only lands in a single orbit O_{X^n}, we are effectively testing whether each element of $(X^n)_{n\geq 1}$ is uniform on the orbit it lands in.

Adding a sequence of statistics $(S_n)_{n\geq 1}$ we can analogously define sequential invariance through this sequence of statistics.

Example 7 (i.i.d. invariant random variables)**.** *Perhaps the simplest setting is when we observe i.i.d. random variables $Y_1, Y_2, \dots$ each in some sample space $\mathcal{Y}$. Suppose we are interested in testing whether their shared distribution is invariant under a group $\mathcal{G}_\dagger$. To fit this into our framework, we can simply choose, $X^n = Y_1, \dots, Y_n$, $\mathcal{X}^n = (\mathcal{Y})^n$ and $\mathcal{G}_n = (\mathcal{G}_\dagger)^n$ for all n. This problem is studied by Chiu and Bloem-Reddy [2023], in a non-sequential setting.*

Example 8 (Sequential exchangeability and i.i.d.)**.** *Suppose that $X^n = Y_1, \dots, Y_n$ for each n. Let us choose $\mathcal{G}_n = \mathfrak{P}_n$ as the group of permutations on n elements. If $(X^n)_{n\geq 1}$ is invariant under $(\mathfrak{P}_n)_{n\geq 1}$, then we say that $(X^n)_{n\geq 1}$ is sequentially exchangeable. This is often also called exchangeability, but we use sequential exchangeability to distinguish it from other forms of exchangeability. With regards to the Type I or post-hoc error Type I, testing sequential exchangeability is equivalent to testing whether the sequence is i.i.d.. This is because the convex hulls of the distributions of i.i.d. and exchangeable sequences coincide by de Finetti's Theorem, and these errors are closed under convex combinations [Vovk, 2021, Ramdas et al., 2022b].*

Example 9 (Within-batch exchangeability)**.** *Suppose we sequentially observe potentially unequally sized batches of data $Y_1, Y_2, \dots,$ where each Y_i is exchangeable, $i = 1, 2 \dots$. We can choose $\mathcal{G}_n = \mathfrak{P}^1 \times \mathfrak{P}^2 \times \cdots \times \mathfrak{P}^n$, where $\mathfrak{P}^i$ is the group of permutations acting on the*

[1] This latent random variable is introduced for ease of exposition and it needs not be modelled or 'exist'.

[2] This is well-defined if and only if $\text{proj}_{\mathcal{X}^n}(x^1) = \text{proj}_{\mathcal{X}^n}(x^2) \implies \text{proj}_{\mathcal{X}^n}(Gx^1) = \text{proj}_{\mathcal{X}^n}(Gx^2)$ for all $G \in \mathcal{G}_n$ and $x^1, x^2 \in \mathcal{X}$ (see e.g. Theorem 2.4 in Eaton 1989).

batch Y_i. Defining $X^n = Y_1, \dots Y_n$, within-batch exchangeability can be viewed as invariance of $(X^n)_{n\geq 1}$ under this group $(\mathcal{G}_n)_{n\geq 1}$.

If we view the elements of a batch as individual observations, then within-batch exchangeability is weaker than sequential exchangeability of the individual observations: we exclude permutations that swap observations across batches. Specifically, the groups we consider here are subgroups of the permutations on the set of the individual observations.

3.2 Post-hoc anytime valid p-process for group invariance

In this section, we construct an anytime valid p-process based on martingales. Recall that in the non-sequential setting, testing invariance comes down to testing whether the random variable has a conditional distribution that is uniform on its orbit. In the sequential setup, we must additionally adapt to the filtration $(\sigma(X^n))_{n\geq 1}$. Specifically, when using a martingale we are effectively testing whether X^n is uniform on O_{X^n} given X^{n-1} and O_{X^n} at each step n. This means that we are effectively testing under the filtration $\sigma(X^1, O_{X^2}) \subseteq \sigma(X^2, O_{X^3}) \subseteq \cdots$.

To adapt to this effective filtration, we characterize the conditional distribution of X^n given X^{n-1} and O_{X^n} in Proposition 1. A key ingredient is the subgroup that stabilizes the past data

$$\mathcal{K}_n(X^{n-1}) = \{G \in \mathcal{G}_n : GX^{n-1} = X^{n-1}\},$$

for $n \geq 2$ and $\mathcal{K}_1 = \mathcal{G}_1$. In Appendix A, we show that this is indeed a compact subgroup of $\mathcal{G}_n$, and include a proof of a more general result.

Proposition 1. *Suppose that $X^n = x^n$. Let $\overline{K}_n$ be uniform on $\mathcal{K}_n(x^{n-1})$, $x^{n-1} = \mathrm{proj}_{\mathcal{X}^{n-1}}(x^n)$. Then, conditional distribution of X^n given X^{n-1} and O_{X^n} is equal to the law of $\overline{K}_n x^n$. That is, the distribution is uniform on the orbit of x^n under $\mathcal{K}_n(x^{n-1})$.*

Having defined these subgroups, we can derive a post-hoc p-process. For each n, let $T_n : \mathcal{X}^n \to \mathbb{R}$ be a non-negative test statistic that is designed to be 'large' under the alternative. This sequence of test statistics $(T_n)_{n\geq 1}$ is allowed to depend on $\sigma(X^{n-1}, O_{X^n})$.

We introduce the following p-process for group invariance, with respect to the filtration $(\sigma(X^n))_{n\geq 1}$:

$$\mathfrak{p}_n(X^n) = \prod_{i=1}^{n} \frac{\mathbb{E}_{\overline{K}_i} T_i(\overline{K}_i X^i)}{T_i(X^i)}.$$

Theorem 3 shows that this is indeed an exact post-hoc p-process. This p-process can be interpreted as a product of p-values for $\mathcal{K}_n(X^{n-1})$ invariance. That is, we are effectively testing whether $(X^n)_{n\geq 1}$ is $(\mathcal{K}_n(X^{n-1}))_{n\geq 1}$ invariant through $(T_n)_{n\geq 1}$.

Theorem 3. *$(\mathfrak{p}_n)_{n\geq 1}$ is an exact post-hoc p-process for $(\mathcal{G}_n)_{n\geq 1}$ invariance through $(T_n)_{n\geq 1}$ with respect to the filtration $(\sigma(X^n))_{n\geq 1}$.*

Example 10 (Sequential sphericity)**.** *Suppose that $\mathcal{X}^n = \mathbb{R}^n$ so that X^n is a random n-vector for all n. Let $\mathcal{O}_n$ be the collection of $n \times n$ orthonormal matrices. Then, X^n is said to be spherically distributed if it is invariant under $\mathcal{O}_n$. Let us consider sequential sphericity,* *where $(X^n)_{n\geq 1}$ is invariant under matrix multiplication by the orthonormal matrices in $(\mathcal{O}_n)_{n\geq 1}$.*

In this example, the orbit O_{X^n} is the hypersphere in n dimensions that contains X^n. As a consequence, the effective filtration reveals the previous observations X^{n-1} and the length of X^n. Together, these determine X^n up to the sign of its final element. As a result, $\mathcal{K}_n$ contains two elements: $\mathrm{diag}(1,\ldots,1,1)$ and $\mathrm{diag}(1,\ldots,1,-1)$, which flip the sign of the final element. This is equivalent to testing whether X^n is sequentially invariant under sign-flips.

Example 11 (Post-hoc p-process for sequential exchangeability)**.** *Continuing from Example 8, suppose we sequentially observe $X^n = Y_1, Y_2, \ldots, Y_n$ that are exchangeable.*

Here, it turns out that X^n is degenerate conditional on $\sigma(X^{n-1}, O_{X^n})$. In particular, $X^{n-1} = Y_1, Y_2, \ldots, Y_{n-1}$ and O_{X^n} equals the multiset $\{Y_1, \ldots Y_n\}$. Hence, Y^n is simply the value in O_{X^n} that is not accounted for in X^{n-1}. As a consequence, the conditional distribution X^n given X^{n-1} and O_{X^n} is degenerate. Assuming the realizations are distinct, this means $\mathcal{K}_n$ only contains the identity element for each n.

A consequence is that it is impossible to sequentially test sequential exchangeability with a (super)martingale under the filtration $(\sigma(X^n))_{n\geq 1}$, as previously observed by Vovk [2021] and Ramdas et al. [2022a].

Example 12 (Post-hoc p-process for within-batch exchangeability)**.** *Continuing from Example 9, let us again consider $X^n = Y_1, \ldots Y_n$, where each Y_i is an exchangeable batch of data. Let us assume the realizations are distinct in each batch. Then, $\mathcal{K}_n(X^{n-1}) = \{I^1\} \times \{I^2\} \times \cdots \times \{I^{n-1}\} \times \mathfrak{P}^n$, where I^i denotes the identity permutation acting on the ith batch. That is, the conditional distribution of X^n is uniform on the final batch. Interestingly, the stabilizer $\mathcal{K}_n(X^{n-1})$ does not depend on X^{n-1},*

As discussed in Example 9, sequential exchangeability implies within-batch exchangeability. This means rejecting within-batch exchangeability also rejects sequential exchangeability. As a result, we can construct a sequential test for sequential exchangeability by merging observations into batches. This of course impoverishes the filtration, since we only look at the data after a batch has arrived. The size of a batch is allowed to depend on the pre-batch data. Generalizing this reasoning is the topic of Section 4.

4 Modifying the filtration

In the previous section, we considered sequential testing where the number of observations could depend on the filtration $(\sigma(X^n))_{n\geq 1}$ which reveals the full data X^n at each step n. While this allows for the number of observations to depend arbitrarily on the data, this can come at a great cost of statistical power, as highlighted in Example 11. This is because we are effectively using the filtration $(\sigma(X^n, O_{X^{n+1}}))_{n\geq 1}$, since testing invariance is testing whether X^n is uniform on O_{X^n}, conditional on O_{X^n}. This effective filtration can be so rich that it reveals or almost reveals the next observation X^{n+1}.

In many practical situations, we may not look at all the data to decide the number of observations, but only consider certain statistics of the data. Such statistics produce less informative filtrations. The topic of this section will be to sequentially test group invariance under such impoverished filtrations.

We treat two approaches, where the second approach generalizes the first. In the first approach, we transform X^n through a statistic $H_n : \mathcal{X}^n \to \mathcal{H}^n$ into $H^n(X^n)$, and induce a group action on the codomain $\mathcal{H}^n$. The problem then reduces to testing the induced invariance of $H_n(X^n)$. Can then directly apply the methodology in Section 3.2 to this reduced problem. The second approach relies on invariance through a statistic, as specified in Definition 2. This second approach nests the methodology of Vovk [2021] and its recent generalizations by Lardy and Pérez-Ortiz [2024]. The key improvement is that our approach does not require to go through ranks, and more abstractly does not require the distribution of the statistic to be the same on every orbit.

Remark 1. *Ramdas et al. [2022b] offer an alternative approach that uses a post-hoc p-process which is not the reciprocal of a (super)martingale, in the context of binary and d-ary data. Unfortunately, it is not clear how to generalize their approach in a useful manner. Moreover, Vovk [2023a] argues that for sequential exchangeability this approach only works for "toy situations" such as binary data. In addition, post-hoc p-processes that do not rely on martingales are generally less flexible [Ramdas et al., 2022b]. A consequence is that we cannot freely change the test statistic after every observation, based on the past data.*

4.1 First approach: invariance of a statistic

On an abstract level, we reduce the problem from testing $\mathcal{G}_n$ invariance of X^n, to testing $\mathcal{F}_n$ invariance of a statistic $H_n(X^n)$, where $\mathcal{F}_n$ is a subgroup of $\mathcal{G}_n$. This allows us to disregard the original data X^n and group $\mathcal{G}_n$, and behave as if we only observe the statistic $H_n(X^n)$ and subgroup $\mathcal{F}_n$. We can then construct a p-process for this reduced problem as in Section 3.2. As $(\mathcal{G}_n)_{n\geq 1}$ of $(X^n)_{n\geq 1}$ implies $(\mathcal{F}_n)_{n\geq 1}$ invariance of $(H_n(X^n))_{n\geq 1}$, the resulting p-process is post-hoc anytime valid for the original problem.

Specifically, suppose we have another space $\mathcal{H}$, and a nested sequence of subsets $(\mathcal{H}^n)_{n\geq 1}$. Moreover, assume we have a sequence of continuous projection maps $\text{proj}_{\mathcal{H}^n} : \mathcal{H} \to \mathcal{H}^n$. We assume that our statistics map into these subsets $H_n : \mathcal{X}^n \to \mathcal{H}^n$. Moreover, we assume that we have a compact subgroup $\mathcal{F}_n$ of $\mathcal{G}_n$ such that H_n induces a group action on $\mathcal{H}^n$ through the group action of $\mathcal{F}_n$ on $\mathcal{X}^n$: $H_n(F_n X^n) = F_n(H_n(X^n))$. This ensures $\mathcal{F}_n$ partitions $\mathcal{H}^n$ into orbits $\mathcal{H}^n/\mathcal{F}_n$. Under these assumptions, the problem reduces to testing $\mathcal{F}_n$ invariance of $(H_n(X^n))_{n\geq 1}$ under the reduced filtration $(\sigma(H_n(X^n)))_{n\geq 1}$. The effective filtration becomes $(\sigma(H_n(X^n), O_{H_n(X^n)}))_{n\geq 1}$ where $O_{H_n(X^n)}$ is the orbit of $H_n(X^n)$ under $\mathcal{F}_n$ in $\mathcal{H}^n$.

In Example 13, we provide an illustration of this approach. In particular, we show how we can test sequential exchangeability by reducing the filtration in a way that yields the within-batch exchangeability approach discussed in Example 9 and Example 12.

Example 13. *Suppose $X^n = Y_1, \dots, Y_n$, and that $(X^n)_{n\geq 1}$ is sequentially exchangeable. We now consider a statistic H_n that effectively censors $(X^n)_{n\geq 1}$ so that we only observe it in 'batches'. Let $b_1, b_2, \dots$ denote the observation numbers at which a batch is completed, and B_n the number of completed batches at time n. Then, we define the statistic to equal the most recently arrived batch of data $H_n(X^n) = X^{b_i}$ and similarly define the reduced sample space $\mathcal{H}^n = \mathcal{X}^{b_i}$, for all $b_i \leq n < b_{i+1}$, $i < B_n$. To induce a group action on $\mathcal{H}^n$, we pass from the group of all permutations $\mathfrak{P}_n$ to its subgroup $\mathcal{F}_n = \mathfrak{P}^1 \times \mathfrak{P}^2 \times \cdots \times \mathfrak{P}^{B_n} \times I$,* *where $\mathfrak{P}^i$ permutes the observations in the ith batch of data, and I acts as the identity on the yet to be completed batch. It remains to verify that this indeed induces a group action. For this, we need to check whether $H_n(x_1^n) = H_n(x_2^n)$ implies $H_n(Fx_1^n) = H_n(Fx_2^n)$ for all $F \in \mathcal{F}_n$ and $x_1^n, x_2^n \in \mathcal{X}^n$. This is equivalent to checking whether $x_1^{b_i} = x_2^{b_i}$ implies $H_n(Fx_1^n) = H_n(Fx_2^n)$, where $b_i \leq n < b_{i+1}$, $i < B_n$. This is indeed satisfied, as F only acts on the already completed batches.*

4.2 Second approach: invariance through a statistic

We now generalize the approach in Section 4.1. In particular, we will not induce a group action on the codomain of a statistic. Instead, we rely on invariance through a statistic, as introduced in Definition 2. This yields our most general construction of a post-hoc valid p-process.

In particular, the strategy is to select statistics $(S_n)_{n\geq 1}$ and compact subgroups $(\mathcal{F}_n)_{n\geq 1}$ so that X^n is $\mathcal{F}_n$ invariant through S_n, conditional on $S_{n-1}(X^{n-1})$. This can happen even though X^n is not $\mathcal{F}_n$ invariant conditional on $S_{n-1}(X^{n-1})$.

Under this assumption, we can construct a post-hoc p-process for $(\mathcal{G}_{n\geq 1})$ with respect to the filtration $(\sigma(S_n(X^n)))_{n\geq 1}$:

$$\mathfrak{p}_n(X^n) = \prod_{i=1}^{n} \frac{\mathbb{E}_{\overline{F}_i} T_i(S_i(\overline{F}_i X^i))}{T_i(S_i(X^i))}, \tag{6}$$

where $T_i : \mathcal{X}^i \to \mathbb{R}_+$ may depend on the filtration, $i \leq n$. Theorem 4 shows this is indeed a post-hoc valid p-process, and even exactly so. A proof, as well as other proofs of the results here can be found in the Appendix.

Theorem 4. *Suppose X^n is $\mathcal{F}_n$ invariant through S_n, conditional on $S_{n-1}(X^{n-1})$ for all n. Then, $(\mathfrak{p}_n(X^n))_{n\geq 1}$ as in (6) is an exact post-hoc valid p-process.*

In Proposition 2, we provide sufficient conditions so that $S_n(X^n)$ is independent from $S_{n-1}(X^{n-1})$. Under these conditions, we can drop the conditioning on $S_{n-1}(X^{n-1})$ in Theorem 4 and we need only check whether X^n is $\mathcal{F}_n$ invariant through S_n for each n. Here, the key condition is that S_n is $\mathcal{F}_{n-1}$ invariant. Indeed, the assumption that $\mathcal{F}_{n-1}$ is a subgroup of $\mathcal{F}_n$ is only used to ensure the group action of $\mathcal{F}_{n-1}$ on $\mathcal{X}^n$ is well-defined.

Proposition 2. *Suppose $\mathcal{F}_{n-1}$ is a subgroup of $\mathcal{F}_n$, and that S_n is a $\mathcal{F}_{n-1}$ invariant function. Then $S_n(X^n)$ is independent of X^{n-1}.*

Similar assumptions as in Proposition 2 are made by Lardy and Pérez-Ortiz [2024], who generalize the work of Vovk [2021]. They make two additional assumptions: $\mathcal{F}_n = \mathcal{G}_n$ for all n, and S_n is a certain smoothed-rank-type statistic. Together, their additional assumptions implicitly ensure that $S_n(X^n)$ has the same distribution on every orbit of $\mathcal{X}^n$ under $\mathcal{G}_n$. That is, the distribution of $S_n(X^n)$ is orbit independent, as in Section 2.2. This implies we need not condition on the orbit of X^n under $\mathcal{F}_n$ to construct the p-process.

In Theorem 5, we provide conditions that more tightly nest the assumptions of Vovk [2021] and Lardy and Pérez-Ortiz [2024]. The main result of Lardy and Pérez-Ortiz [2024] is recovered by choosing S_n as a specific smoothed-rank-type statistic, which ensures the fourth condition holds. The approach of Vovk [2021] is recovered by additionally reducing to sequential exchangeability. In examples 14 and 15, we detail how their methodology fits into Theorem 5.

Theorem 5. *Suppose the following conditions hold:*

- $\mathcal{F}_{n-1}$ *is a subgroup of* $\mathcal{F}_n$*,*
- X^n *is* $\mathcal{F}_n$ *invariant through* S_n*,*
- S_n *is a* $\mathcal{F}_{n-1}$ *invariant function,*
- $S_n(X^n)$ *has the same distribution on every orbit of* $\mathcal{X}^n$ *under* $\mathcal{F}_n$*.*

Then, the conditional distribution of $S_n(X^n)$ *given* $\sigma(S_{n-1}(X^{n-1}), O_{X^n})$ *is equal to the distribution of* $S_n(\overline{F}_n X^n)$*, where* $\overline{F}_n \sim \mathrm{Unif}(\mathcal{F}_n)$*.*

Remark 2. *It is also possible to add external randomization into the statistic* S_n*, which is used by Lardy and Pérez-Ortiz [2024]. This can be used, for example, to break ties or get rid of other discretenesses in order to ensure* $S_n(X^n)$ *has the same distribution on every orbit. We use this in Example 15.*

Example 14 (Sequential ranks)**.** *In this example, we show how the methodology of Vovk [2021] for sequential exchangeability using ranks fits into our framework.*

Suppose we are in the sequential exchangeability setting as in Example 8 and 11. That is, we have data $X^n = Y_1, Y_2, \ldots, Y_n$*, and* X^n *is exchangeable for each* n*. Suppose that we additionally assume that* $Y_1, Y_2, \ldots$ *can be deterministically ordered. This holds, for example, if they are real-valued and have no ties. We indeed have that* $\mathcal{G}_{n-1}$ *is a subgroup of* $\mathcal{G}_n$ *here, as the permutations on* $n-1$ *elements are included in the permutations on* n *elements. Moreover,* X^n *is* $\mathcal{G}_n$ *invariant, so that it is* $\mathcal{G}_n$ *invariant through any statistic.*

We now consider the sequential rank statistic $S_n = \mathrm{lastRank}_n : \mathcal{X}^n \to \{1, \ldots, n\}$*. For a given input* X^n*, it returns the rank of its* n*th element* Y_n *among the preceding elements* $Y_1, \ldots, Y_{n-1}$*. For example, if* $X^n = 7, 3, 1, 4$*, then* $\mathrm{lastRank}_n(X^n) = 3$*, as* 4 *is the 3rd smallest number.*

It remains to verify the final two conditions of Theorem 5. First, as X^n *is exchangeable, the rank of its final element is uniformly distributed on* $\{1, \ldots, n\}$*, regardless of the distribution of* X^n*. Next,* S_n *is indeed a* $\mathcal{G}_{n-1}$ *invariant function: permuting its first* $n-1$ *elements has no impact on the rank of the final element.*

This means we can apply Theorem 5. As a consequence, for any test statistic $T_n : \{1, \ldots, n\} \to \mathbb{R}_+$*, we have that a post-hoc anytime valid p-process for sequential exchangeability is given by*

$$\mathfrak{p}_n(X^n) = \prod_{i=1}^{n} \mathbb{E}_{\overline{P}_i} T_i(\mathrm{lastRank}_i(\overline{P}_i X^i)) / T_i(\mathrm{lastRank}_i(X^i)),$$

where $\overline{P}_n$ *is uniform on the group of permutations on* n *elements.*

Example 15 (Sequential sphericity)**.** *Lardy and Pérez-Ortiz [2024] generalize Example 14 to other types of invariance. In this example, we show how their methodology fits into our framework, by illustrating it on sequential sphericity as in Example 10. As will be shown* *below, the key difference is that our framework does not necessitate the use of (smoothed) ranks.*

We sequentially observe $Y_1, Y_2, \ldots,$ and assume that $X^n = (Y_1, \ldots Y_n)'$ is a spherical random n-vector for each n. Lardy and Pérez-Ortiz [2024] consider the 'smoothed rank' statistic $R_n : X^n \mapsto \mathbb{P}_{\overline{G}_n}[m_n(e_n'\overline{G}_n X^n) > m_n(e_n' X^n)] + \overline{u}\mathbb{P}_{\overline{G}_n}[m_n(e_n'\overline{G}_n X^n) = m_n(e_n' X^n)]$, where m_n is some function that can depend on the previous data, $\overline{u} \sim \mathrm{Unif}[0,1]$ and e_n is the nth canonical basis vector.

We proceed by checking the conditions of Theorem 5. Here, the group $\mathcal{G}_n$ is the orthogonal group in dimension n, which contains the orthogonal group in dimension $n-1$ as a subgroup. Next, X^n is $\mathcal{G}_n$ invariant, so it is indeed $\mathcal{G}_n$ invariant through R_n. Furthermore, $e_n'G = e_n$ for all $G \in \mathcal{G}_{n-1}$ as $\mathcal{G}_{n-1}$ only acts on the first $(n-1)$ elements. As a result, R_n is $\mathcal{G}_{n-1}$ invariant. Finally, as X^n is $\mathcal{G}_n$ invariant, $R^n(X^n) \sim \mathrm{Unif}[0,1]$ regardless of the orbit of X^n, as shown in Theorem 1.

Hence, all the conditions for Theorem 5 are satisfied with the statistics $(R_n)_{n\geq 1}$ and groups $(\mathcal{G}_n)_{n\geq 1}$. Lardy and Pérez-Ortiz [2024] then propose the post-hoc p-process

$$\mathfrak{p}_n^*(X^n) = \prod_{i=1}^{n} \mathbb{E}_{\overline{u}} T_i^*(\overline{u}) / T_i^*(R_i(X^i)),$$

where $T_i^* : [0,1] \to \mathbb{R}_+$ can depend on the past data in the filtration, $\overline{u} \sim \mathrm{Unif}[0,1]$ and $\mathbb{E}_{\overline{u}} T_i^*(\overline{u}) < \infty$. The $\overline{u}$ appears, since $\overline{u} \stackrel{d}{=} R^n(\overline{G}_n X^n)$ for all n. Indeed, $\mathbb{E}_{\overline{u}} T_i^*(\overline{u}) = \mathbb{E}_{\overline{G}_n} T_i^*(\overline{G}_n X^n)$.

To show that our methodology is more flexible, we now construct a post-hoc p-process based on another statistic. Specifically, let us consider the statistic $S_n : \mathcal{X}^n \to [-1,1]$, defined as $S_n(X^n) = e_n' X^n / \|X^n\|_2$, where e_n is the nth canonical basis vector. Is is easy to verify that $H_n(X^n) = S_1(X^1), \ldots, S_n(X^n)$ induces a filtration, so we proceed with checking the conditions of Theorem 5.

The first two conditions of Theorem 5 are easily verified. It remains to check the final two conditions. First, $e_n'G = e_n$ for all $G \in \mathcal{G}_{n-1}$ and $\|GX^n\|_2 = \|X^n\|_2$ for all $G \in \mathcal{G}_n \supseteq \mathcal{G}_{n-1}$. As a result, $S_n(GX^n) = e_n' G X^n / \|GX^n\|_2 = e_n' X^n / \|X^n\|_2$, so S_n is indeed $\mathcal{G}_{n-1}$ invariant. Second, as X^n is $\mathcal{G}_n$ invariant, $X^n / \|X^n\|_2$ is uniform on the unit hypersphere in dimension n regardless of the orbit of X^n. As a consequence, $S_n(X^n) = e_n' X^n / \|X^n\|_2$ has the same distribution for every orbit.

As all its conditions are satisfied, we can apply Theorem 5. As a consequence, for any test statistic $T_n : [-1,1] \to \mathbb{R}_+$, we have that a post-hoc anytime valid p-process for sequential sphericity is given by

$$\mathfrak{p}_n(X^n) = \prod_{i=1}^{n} \mathbb{E}_{\overline{G}_i} T_i(S_i(\overline{G}_i X^i)) / T_i(S_i(X^i)),$$

where $\overline{G}_n$ is uniform on the orthogonal group in n dimensions.

5 Likelihood ratios for group invariance

The post-hoc p-values we derived are highly flexibly in the choice of the test statistic. Unfortunately, this freedom also comes with the responsibility to select the test statistic appropriately. For testing a simple null hypothesis against a simple alternative, likelihood ratios are exact post-hoc p-values with attractive power-like properties [Shafer, 2021, Koolen and Grünwald, 2022, Grünwald et al., 2023, Ramdas et al., 2022a]. In particular, for a simple null $H_0 = \{\mathbb{P}_0\}$ versus simple alternative $H_1 = \{\mathbb{P}_1\}$, the post-hoc p-value that maximizes $\mathbb{E}^{\mathbb{P}_1} \log 1/p$ is the reciprocal of the likelihood ratio: $1/p = d\mathbb{P}_1/d\mathbb{P}_0$. Such a post-hoc p-value is also called log-optimal or growth-rate optimal (GRO).

In additional, Koning [2024] notes that the reciprocal of such a post-hoc p-value is also post-hoc valid under the alternative. This means that it can be interpreted both as evidence against the null and against the alternative.

For these reasons, we derive likelihood ratio statistics for group invariance, which are well-specified even though the null hypothesis is highly composite. This is because invariance is equivalent to uniformity on every orbit, so that on each orbit we have a simple null hypothesis.

It turns out that these likelihood ratio statistics come in three flavors, where the flavor depends on the space on which we specify our alternative. For the first flavor, we specify an alternative on the entire sample space $\mathcal{Y}$. This type is inspired by a proof strategy of Lehmann and Stein [1949], who did not explicitly construct the likelihood ratio statistic but only derived a test that is equivalent to the likelihood ratio test (see Remark 3). For the second flavor, we do not specify an alternative on $\mathcal{Y}$, but we specify an alternative on every orbit. Although the orbits partition the sample space, this strictly generalizes the first flavor, as we need not specify the mixing distribution over the orbits. For the third and final flavor, we specify an alternative on the group $\mathcal{G}$. While we do not directly observe an element on our group, we use a so-called inversion kernel [Kallenberg, 2017] to obtain such an element.

The ideas in this section are illustrated in Section 6 to testing a location shift under Gaussianity against various types of invariances. Moreover, we include a toy example in Appendix B to illustrate the concepts.

5.1 Alternative on the sample space $\mathcal{Y}$

Suppose that $\mathbb{P}_{\mathcal{Y}}$ is our alternative on $\mathcal{Y}$, dominated by some measure λ, so that we can define the density $d\mathbb{P}_{\mathcal{Y}}/d\lambda$. A likelihood ratio statistic for testing this alternative against $\mathcal{G}$ invariance is presented in Theorem 6. Its reciprocal is a post-hoc p-value for $\mathcal{G}$ invariance. A proof is presented in Appendix D.7.

Theorem 6. *Let $\overline{G}$ be uniform $\mathcal{G}$ and assume that $0 < \mathbb{E}_{\overline{G}} d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}y) < \infty$ for all $y \in \mathcal{Y}$. Then, the statistic*

$$\frac{d\mathbb{P}_{\mathcal{Y}}/d\lambda(y)}{\mathbb{E}_{\overline{G}} d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}y)} \tag{7}$$

is a likelihood ratio statistic between $\mathcal{G}$ invariance and $\mathbb{P}_{\mathcal{Y}}$.

Note that this coincides with our post-hoc p-value with the statistic $T = d\mathbb{P}_{\mathcal{Y}}/d\lambda$. The result also holds if T is merely proportional to a density, as the proportionality constant drops out. Hence, for a statistic T that is proportional to a density on $\mathcal{Y}$, we can interpret the resulting e-value as a likelihood ratio statistic against this density.

Remark 3. *The proof of Theorem 6 mimics the proof strategy of Theorem 2 and 2′ in Lehmann and Stein [1949]. Interestingly, they do not explicitly derive this likelihood ratio statistic. Instead, they show that the group invariance test based on the statistic $d\mathbb{P}_{\mathcal{Y}}/d\lambda$,*

$$d\mathbb{P}_{\mathcal{Y}}/d\lambda(y) > q_{\alpha}^{\overline{G}}\left(d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}y)\right), \tag{8}$$

is equivalent to a likelihood ratio test and hence uniformly most powerful for testing $\mathcal{G}$ invariance against $\mathbb{P}_{\mathcal{Y}}$ by the Neyman-Pearson lemma. We suspect that they did not explicitly compute the likelihood ratio statistic itself as the test in (8) *is a much more efficient representation of the likelihood ratio test. Indeed, it is equivalent to the test*

$$\frac{d\mathbb{P}_{\mathcal{Y}}/d\lambda(y)}{\mathbb{E}_{\overline{G}}d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}y)} > q_{\alpha}^{\overline{G}}\left(\frac{d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}y)}{\mathbb{E}_{\overline{G}_2}d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}_2y)}\right),$$

but does not require the computation of the normalization constants, as they drop out.

5.2 Alternative on the orbits

In the previous section we defined an alternative on the entire sample space $\mathcal{Y}$. However, a careful inspection of the proof shows that we only compare the likelihood of y to the likelihood of other values on its orbit. This implies that the likelihood ratio statistic only uses the conditional distributions on the orbits, and 'discards' the mixing distribution over the orbits.

As a consequence, we can also define a likelihood ratio statistic by specifying an alternative on every orbit. Specifically, let us choose a distribution $\mathbb{P}_z$ on each orbit O_z, that is absolutely continuous with respect to λ_z: the unique $\mathcal{G}$ invariant (uniform) distribution on O_z, $z \in [\mathcal{Y}]$. Then, we can construct a likelihood ratio statistic for each orbit:

$$\frac{d\mathbb{P}_z/d\lambda_z(y)}{\mathbb{E}_{\overline{G}}d\mathbb{P}_z/d\lambda_z(\overline{G}y)} = d\mathbb{P}_z/d\lambda_z(y),$$

where $z \in [\mathcal{Y}]$, $y \in O_z$, and the equality follows from the fact that the numerator is equal to 1 on each orbit, as it is a density on the orbit. This simplification is not possible in (7), as $dP_{\mathcal{Y}}/d\lambda$ is only a density on $\mathcal{Y}$, which needs not integrate to 1 on each orbit.

With this observation, we can actually slightly strengthen the main result (Theorem 2 and 2′) of Lehmann and Stein [1949], by discarding the mixture distribution.

Theorem 7. *Suppose that T is an strictly increasing transformation of h on $\mathcal{Y}$. Let us consider the group invariance test based on the statistic T, which rejects if*

$$T(y) > q_{\alpha}^{\overline{G}}(T(\overline{G}y)),$$

and with some appropriate probability in case of equality. Then, this test is not just uniformly most powerful against the density h (as shown by Lehmann and Stein [1949]), but against the composite alternative that consists of all distributions with the same conditional distributions on the orbits as h.

5.3 Alternative on the group $\mathcal{G}$ and inversion kernels

In this section, we define a likelihood ratio statistic based on an alternative $\mathbb{P}^1_{\mathcal{G}}$ on the group $\mathcal{G}$. Let $\mathbb{P}^0_{\mathcal{G}}$ denote the unique Haar measure on the group, so that we can define a likelihood ratio as

$$d\mathbb{P}^1_{\mathcal{G}}/d\mathbb{P}^0_{\mathcal{G}}(G).$$

Unfortunately, this likelihood ratio is infeasible, as we do not directly observe an element of the group but only an element in our sample space $\mathcal{Y}$.[3] However, this can be resolved with use of a so-called inversion kernel (see Chapter 7 of Kallenberg [2017]), which was first introduced in the context of group invariance testing by Chiu and Bloem-Reddy [2023].

To start, let us assume that $\mathcal{G}$ acts *freely* on $\mathcal{Y}$. This means that $Gy = y$ implies $G = I$. In this case, we can uniquely define a so-called *inversion kernel* $\gamma : \mathcal{Y} \to \mathcal{G}$ that takes an element y and returns the element G that carries the representative element $[y]$ on the orbit of y to y. That is, $\gamma(y)[y] = y$. For example, if no element in a vector $x \in \mathbb{R}^d$ has duplicated elements, then the group of permutations acts freely on it: any non-identity permutation of x would yield a different vector.

If the group action is not free, then there may exist multiple elements in $\mathcal{G}$ that carry $[y]$ to y, so that $\gamma(y)$ is not uniquely defined. For the non-free setting, we overload the notation of γ so that $\gamma(y)$ is uniformly drawn from the elements in $\mathcal{G}$ that carry $[y]$ to y, which is well-defined by Theorem 7.14 in Kallenberg [2017]. This gives us $\gamma(y)[y] = y$ almost surely. Appendix B contains a concrete illustration of a setting where γ is randomized in this manner, and an intuition of why it is possible to construct a uniform draw from such elements.

We can also use γ to obtain an alternative characterization of $\mathcal{G}$ invariance of a random variable:

$$\gamma(Y) \overset{d}{=} \overline{G}, \tag{9}$$

where $\overline{G}$ is uniformly (Haar) distributed on $\mathcal{G}$ (see e.g. Chiu and Bloem-Reddy 2023).

Using this map γ, we can define the (randomized) likelihood statistic

$$d\mathbb{P}^1_{\mathcal{G}}/d\mathbb{P}^0_{\mathcal{G}}(\gamma(y)). \tag{10}$$

Alternatively, we can induce a distribution on $\mathcal{G}$ through a distribution on our sample space. In particular, we can start by defining an alternative $\mathbb{P}_{\mathcal{Y}}$ on $\mathcal{Y}$ and let $\mathbb{P}^{\mathcal{Y}}_{\mathcal{G}}$ denote the distribution of $\gamma(\widetilde{Y})$ if $\widetilde{Y} \sim \mathbb{P}_{\mathcal{Y}}$. Then, we can also consider the likelihood ratio statistic

$$d\mathbb{P}^{\mathcal{Y}}_{\mathcal{G}}/d\mathbb{P}^0_{\mathcal{G}}(\gamma(y)),$$

which can be interpreted as testing against the non-invariance expressed by $\widetilde{Y}$ on $\mathcal{G}$.

6 Illustration: LRs for invariance vs Gaussianity

In this section, we illustrate our likelihood ratios to test for invariance under a group of orthonormal matrices against a normal distribution with a location shift. If we include all

[3]Unless our sample space $\mathcal{Y}$ happens to $\mathcal{G}$.

orthonormal matrices, this yields clean connections to parametric theory and Student's t-test. Moreover, we also consider exchangeability, which reveals an interesting relationship to the softmax function. In addition, we consider sign-symmetry, where we provide a relationship to a well-known post-hoc p-value based on de la Peña [1999].

We start with an exposition of the invariance-based concepts for the orthogonal group $O(d)$ that consists of all orthonormal matrices.

6.1 Sphericity

Suppose that $\mathcal{Y} = \mathbb{R}^d \setminus \{0\}$ and $\mathcal{G} = O(d)$ is the orthogonal group, which can be represented as the collection of all orthonormal matrices. The orbits $O_y = \{z \in \mathcal{Y} \mid z = Gy, \exists G \in \mathcal{G}\}$ of $\mathcal{G}$ in $\mathbb{R}^d$ are the concentric d-dimensional hyperspheres. Each of these hyperspheres can be uniquely identified with their radius $\mu > 0$. To obtain a $\mathcal{Y}$-valued orbit representative, we multiply μ by an arbitrary unit d-vector ι to obtain $\mu\iota$. For example y lies on the orbit O_y that is the d-dimensional hypersphere with radius $\|y\|_2$, and has orbit representative $[y] = \|y\|_2\iota$.

For simplicity, we now first focus on the subgroup $SO(2)$ of $O(2)$, which exactly describes the (orientation-preserving) rotations of the circle, and has the same orbits as $O(2)$. The reason we focus on $SO(2)$, is because its group acts freely on each concentric circle. As a consequence, every element in the group can be uniquely identified with an element on the unit circle S^2. We choose to identify the identity element with ι, and we identify every element of $SO(2)$ with the element on the circle that we obtain if that rotation is applied to ι. We denote the induced group action of S^2 on $\mathcal{Y}$ by $\circ$.

We can then define our kernel inversion map γ as $\gamma(y) = y/\|y\|_2$. To see that γ indeed conforms to its definition, observe that

$$\gamma(y)[y] = [(y/\|y\|_2) \circ \iota]\|y\|_2 = (y/\|y\|_2)\|y\|_2 = y, \tag{11}$$

where the second equality follows from the fact that the action of $(y/\|y\|_2)$ on ι, rotates ι to $y/\|y\|_2$. Invariance of an $\mathcal{Y}$-valued random variable Y under $\mathcal{G}$, also known as sphericity, can then be formulated as '$\gamma(Y)$ is uniform on S^2'.

For $O(2)$ or the general $d > 2$ case, the group action is no longer free on each orbit. As a result there may be multiple group actions that carry $\iota\|y\|_2$ to a point y on the hypersphere. While this may superficially seem like a potentially serious issue, we view $\gamma(y)$ as uniformly drawn from all the 'rotations' that carry $\iota\|y\|_2$ to y. As a result, the only difference is that (11) will now hold almost surely, which suffices for our purposes.

6.2 Likelihood ratio on $\mathcal{Y}$

This section can be seen as a generalization of the example in the final paragraph of Lehmann and Stein [1949], who only consider spherical invariance.

Suppose that $Y \sim \mathcal{N}_d(\mu\iota, I)$ on $\mathbb{R}^d \setminus \{0\}$, $\mu \geq 0$ under the alternative and $\mathcal{G}$ invariance under the null hypothesis. This distribution is spherical if and only if $\mu = 0$. We start by considering $\mathcal{G} = O(d)$. The $\mathcal{Y}$-based likelihood ratio test is given by

$$1/(2\pi)^{d/2} \exp\left\{-\frac{1}{2}\|y - \iota\mu\|_2^2\right\} > q_\alpha^{\overline{G}}\left(1/(2\pi)^{d/2} \exp\left\{-\frac{1}{2}\|\overline{G}y - \iota\mu\|_2^2\right\}\right),$$

where $\overline{G}$ is uniformly distributed on all orthonormal matrices. This is equivalent to

$$-y'y + 2\mu\iota'y - \mu^2 > q_\alpha^{\overline{G}}\left(-y'y + 2\mu\iota'\overline{G}y - \mu^2\right)$$

and

$$\iota'y > q_\alpha^{\overline{G}}\left(\iota'\overline{G}y\right),$$

which is independent of μ and equal to the t-test by Theorem 6 in Koning and Hemerik [2023]. As already shown by Lehmann and Stein [1949], the t-test is uniformly most powerful for testing spherical invariance against $\mathcal{N}_d(\mu\iota, I)$.

Moreover, this test can also be written as

$$\iota'y/\|y\|_2 > q_\alpha^{\overline{G}}\left(\iota'\overline{G}\iota\right),$$

as $q_\alpha^{\overline{G}}\left(\iota'\overline{G}y\right) = q_\alpha^{\overline{G}}\left(\iota'\overline{G}\iota\|y\|_2\right) = \|y\|_2 q_\alpha^{\overline{G}}\left(\iota'\overline{G}\iota\right)$. Then, as the rejection event does not change if we apply a strictly increasing function to both sides, we can even conclude that the t-test is equivalent to any spherical group invariance test based on a test statistic that is increasing in $\iota'y/\|y\|_2$.

A straightforward derivation shows that the likelihood ratio statistic is

$$\exp\left\{\mu y'\iota\right\}/\mathbb{E}_{\overline{G}}\left[\exp\left\{\mu y'\overline{G}\iota\right\}\right]. \tag{12}$$

To obtain the likelihood ratio for other groups $\mathcal{G}$ of orthonormal matrices, we can simply compute the normalization constant in (12) with $\overline{G}$ uniform on the group of interest. This includes the group of permutation matrices for testing exchangeability against normality (see Section 6.5), and the group of sign-flipping matrices for testing symmetry against normality (see Section 6.6). The resulting likelihood ratio test is also uniformly most powerful for testing $\mathcal{G}$ invariance against $\mathcal{N}_d(\mu\iota, I)$. Moreover, the reciprocal of this likelihood ratio is a log-optimal post-hoc p-value for the same problem.

6.3 Likelihood ratio on orbits

The conditional distribution of $Y \sim \mathcal{N}_d(\mu\iota, I)$ on each orbit is proportional to $\exp(\mu\iota'y)$, where y is on the orbit with radius $\|y\|_2$. For $\|y\|_2 = 1$, this coincides with the von Mises-Fisher distribution. Notice that this density is uniform on each orbit if and only if $\mu = 0$, so that the likelihood ratio with respect to sphericity is proportional to $\exp(\mu\iota'y)$, and coincides with the one from previous section:

$$\exp\left\{\mu y'\iota\right\}/\mathbb{E}_{\overline{G}}\left[\exp\left\{\mu y'\overline{G}\iota\right\}\right].$$

Applying our argument from Section 5.2, this implies that the t-test is uniformly most powerful against the composite alternative of all distributions on $\mathcal{Y}$ whose conditional distributions on the orbits of $O(d)$ are strictly increasing transformations of $\exp(\mu\iota'y)$. This generalizes the observation by Lehmann and Stein [1949] who only conclude optimality against $\mathcal{N}(\mu\iota, I)$.

6.4 Likelihood ratio on $\mathcal{G}$

In this section, we reduce ourselves to $d = 2$ and $SO(2)$, so that the group action is free and the group will be easy to represent. If $Y \sim \mathcal{N}_2(\mu\iota, I)$, then $\gamma(Y) = Y/\|Y\|_2$ follows a so-called projected normal distribution $\mathcal{PN}_2(\mu\iota, I)$. Its density with respect to the uniform distribution on S^2 is

$$\frac{\exp\{-\frac{1}{2}\mu^2\}}{2\pi}\left(1 + \mu\iota' v\frac{\Phi(\mu\iota' v)}{\phi(\mu\iota' v)}\right),$$

where $v \in S^2$, Φ is the normal cdf and ϕ the pdf (Presnell et al., 1998; Watson, 1983). For $\mu = 0$, this reduces to $1/2\pi$, so the likelihood ratio with respect to the uniform distribution on S^2 is

$$\exp\{-\tfrac{1}{2}\mu^2\}\left(1 + \mu\iota' v\frac{\Phi(\mu\iota' v)}{\phi(\mu\iota' v)}\right).$$

As a result, the likelihood ratio on $\mathcal{Y}$ is

$$dP^1_{\mathcal{G}}/dP^0_{\mathcal{G}}(\gamma(y)) = \exp\{-\tfrac{1}{2}\mu^2\}\left(1 + \mu\iota'\gamma(y)\frac{\Phi(\mu\iota'\gamma(y))}{\phi(\mu\iota'\gamma(y))}\right) = \exp\{-\tfrac{1}{2}\mu^2\}\left(1 + \mu\iota' y/\|y\|_2\frac{\Phi(\mu\iota' y/\|y\|_2)}{\phi(\mu\iota' y/\|y\|_2)}\right)$$

which is an increasing function in $\iota' y/\|y\|_2$ if $\mu > 0$. As the likelihood ratio is increasing in $\iota' y/\|y\|_2$, the likelihood ratio test is also equivalent to the t-test.

6.5 Permutations and softmax

The likelihood ratio in (12) is strongly related to the softmax function. Indeed, if we choose $\overline{G}$ to be uniform on permutation matrices (which form a subgroup of the orthonormal matrices) and $\iota = (1, 0, \ldots, 0)$ this reduces to

$$\frac{\exp\{\mu y_1\}}{\frac{1}{d}\sum_{i=1}^{d}\exp\{\mu y_i\}}. \tag{13}$$

This is exactly the softmax function with 'inverse temperature' $\mu \geq 0$. Hence, the softmax function can be viewed as a likelihood ratio statistic for testing exchangeability (permutation invariance) against $\mathcal{N}((\mu, 0, \ldots, 0), I)$. More generally, it is the likelihood ratio statistic for testing exchangeability on the orbit $O_Y = \{PY \mid P \in \text{permutations}\}$ against the conditional distribution of $\mathcal{N}((\mu, 0, \ldots, 0), I)$ on O_Y.

Remark 4. *A related post-hoc p-value appears in unpublished early manuscripts of Wang and Ramdas [2022] and Ignatiadis et al. [2023], who consider a 'soft-rank' post-hoc p-value of the type* $\mathfrak{p}_T$ *as in* (5) *with*

$$T(y) = \frac{\exp(\kappa y_1) - \exp(\kappa \min_j y_j)}{\kappa}, \tag{14}$$

under exchangeability, for some inverse temperature $\kappa > 0$.

Interestingly, this 'soft-rank' p-value for $\kappa = \mu$ *is smaller than the softmax p-value* (13) *if and only if the softmax p-value is smaller than 1. In fact, the same holds if we replace* $\exp(\kappa \min_j y_i)$ *by any positive constant c, and the relationship flips if c is negative. For a positive constant c, we would therefore expect the 'soft-rank' p-value to have a higher variability.*

6.6 Testing sign-symmetry

Suppose $\mathcal{Y} = \mathbb{R}$ and $\mathcal{G} = \{-1, 1\}$. Then, invariance of Y under $\mathcal{G}$ is also known as 'symmetry' about 0, defined as $Y \overset{d}{=} -Y$. For testing symmetry against our normal location model with $\iota = 1$, the likelihood ratio is

$$\exp\{\mu\iota' y\}/\mathbb{E}_{\overline{G}} \exp\{\mu\iota'\overline{G}y\} = 2\exp\{\mu y\}/\left[\exp\{\mu y\} + \exp\{-\mu y\}\right],$$

This can be generalized to $\mathcal{Y} = \mathbb{R}^d$ and $\mathcal{G} = \{-1, 1\}^d$ and $\iota = d^{-1/2}(1, \ldots, 1)'$. The likelihood ratio then becomes

$$\exp\{d^{-1/2}\mu\iota' y\}/\mathbb{E}_{\overline{g}} \exp\{d^{-1/2}\mu\overline{g}' y\} = \prod_{i=1}^{d} \exp\{d^{-1/2}\mu y_i\}/\mathbb{E}_{\overline{g}_i} \exp\{d^{-1/2}\mu\overline{g}_i y_i\},$$

where $\overline{g}$ is a d-vector of i.i.d. Bernouilli distributed random variables on $\{-1, 1\}$ with probability .5.

Remark 5. *A related post-hoc p-value can be derived from de la Peña [1999], as the reciprocal of*

$$\exp\{Z - Z^2/2\}.$$

This object can be connected to our likelihood ratio, by simply normalizing it by $\mathbb{E}_{\overline{g}}[\exp\{\overline{g}Z - (\overline{g}Z)^2/2\}]$*:*

$$\begin{aligned}\exp\{Z - Z^2/2\}/\mathbb{E}_{\overline{g}}[\exp\{\overline{g}Z - (\overline{g}Z)^2/2\}] &= 2\exp\{Z - Z^2/2\}/\left[\exp\{-Z - Z^2/2\} + \exp\{Z - Z^2/2\}\right] \\ &= 2\exp\{Z\}/\left[\exp\{-Z\} + \exp\{Z\}\right].\end{aligned}$$

This transformation makes the resulting p-value exactly post-hoc by Theorem 2, so that our p-value for sign-symmetry can be interpreted as an exact post-hoc variant of the de la Peña [1999] post-hoc p-value.

Moreover, Ramdas et al. [2022a] characterize the class of admissible post-hoc p-processes for testing symmetry, and show that the p-process based on de la Peña [1999] is inadmissible. In our simulations, we indeed see that it is indeed (strongly) dominated by ours.

7 Simulations

7.1 Case-control experiment and learning the alternative

In this simulation study, we consider a hypothetical case-control experiment, where units are assigned to either the treated or control set uniformly at random. In each interval of time, we receive the outcomes of a number of treated and control units, where the number of both units is Poisson distributed with parameter $\theta > 0$ with a minimum of 1. The outcomes of the treated units are $\mathcal{N}(a, 1)$-distributed and the outcomes of the controls are $\mathcal{N}(b, 1)$-distributed. The true mean and variance are considered unknown, and are adaptively learned based on the previously arrived data. As a batch of data, we will consider the combined observations of both the treated and control units that arrived in the previous interval of time.

As a result, a batch X_t of n^t outcomes, consisting of n_a^t treated and n_b^t control units, can be represented as

$$X_t \sim \begin{bmatrix} 1_{n_a^t} a \\ 1_{n_b^t} b \end{bmatrix} + \mathcal{N}(0, I),$$

where $1_{n_a^t}$ and $1_{n_b^t}$ denote a vector of n_a^t and n_b^t ones, respectively, where the first n_a^t elements correspond to the treated units, without loss of generality. We would like to base our test statistic on the difference of sample means:

$$\overline{1}_{n^t}' X_t \sim \mathcal{N}\left(a - b, 1/n_a^t + 1/n_b^t\right),$$

where $\overline{1}_{n^t} = (1_{n_a^t}(n_a^t)^{-1}, -1_{n_b^t}(n_b^t)^{-1})$. In particular, we will test the null hypothesis that the elements of a batch X_t are exchangeable and so $a = b$, against the alternative hypothesis that $a > b$.

We use our post-hoc p-process based on the likelihood ratio for testing exchangeability against our current estimate of the Gaussian alternative:

$$p_t = \frac{\mathbb{E}_{\overline{P}} \exp\{(\widehat{a}_{t-1} - \widehat{b}_{t-1})/\widehat{\sigma}^2 \times \overline{1}_{n^t}' \overline{P} X_t\}}{\exp\{(\widehat{a}_{t-1} - \widehat{b}_{t-1})/\widehat{\sigma}^2 \times \overline{1}_{n^t}' X_t\}},$$

where $\widehat{a}_{t-1} - \widehat{b}_{t-1} = \overline{1}_{n^t}' X_{t-1}$ is our treatment estimator at time $t-1$ and $\widehat{\sigma}_{t-1}^2$ is its pooled sample variance estimator. For the first batch, we can either rely on an educated guess, or skip it for inference and only use it for estimating these parameters. We estimate the normalization constant by using 100 permutations drawn uniformly at random with replacement.

For our simulations, we consider the arrival of 40 batches with $\theta = 25$. Without loss of generality, we choose $a = b = 0$ under the null, and $a = .2$ and $b = 0$ under the alternative. To use in the first batch, we choose $\widehat{a}_0 = .2$, $\widehat{b}_0 = 0$ and $\hat{\sigma}_0^2 = 1$.

In Figure 1, we plot the post-hoc p-processes based on 1 000 simulations. The plot on the left features the setting under the null, and the plot on the right the setting under the alternative. For convenience, we plot at each time the line which 5%, 50% and 95% of the p-processes have remained above up until that point. For example, in the right plot, roughly 50% of the p-processes have dipped below .05 at batch 21, so that the power at level α is roughly 50% after 23 batches. As expected, left plot shows that the p-processes stay bounded away from every level under the null, and decrease under the alternative.

7.2 Testing symmetry and a comparison to de la Peña [1999]

In this simulation study, we consider testing sign-symmetric data as in Section 6.6. We compare our post-hoc p-process to the one based on de la Peña [1999], when testing against a simple normal alternative $X_i \sim \mathcal{N}(m, 1)$ with $m = 1$.

We plot 1 000 p-processes of each type in Figure 2. The plot on the left is our likelihood ratio-based post-hoc p-process, whereas the plot on the right uses the post-hoc p-value based on de la Peña [1999]. The figure shows that our LR-based p-processes shrink much faster. This coincides with the observation of Ramdas et al. [2022a] that the post-hoc p-process based on de la Peña [1999] is inadmissible.

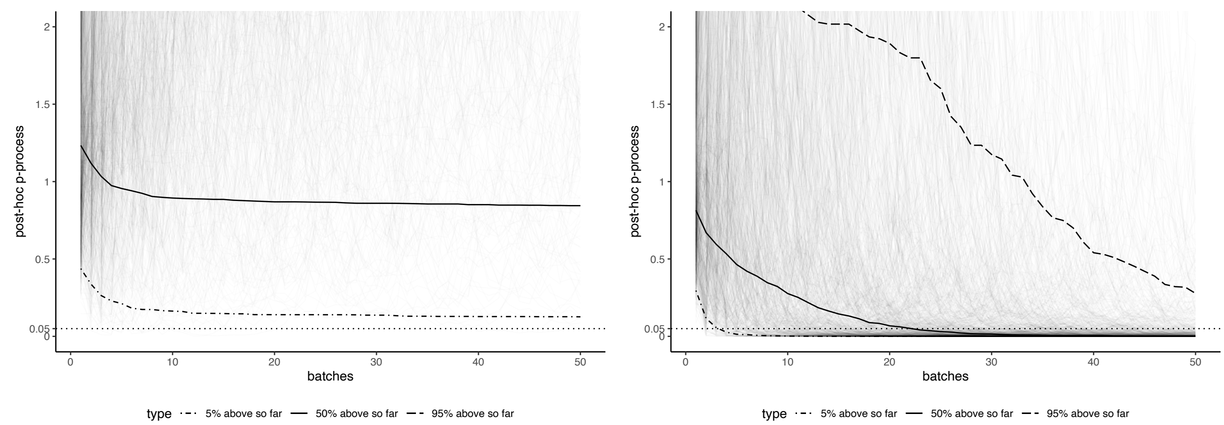

Figure 1: Plots of 1 000 post-hoc p-processes over the number of arrived batches. The highlighted lines are running quantiles: x% of the p-processes have not crossed below the line at the indicated time. The plot on the left is under the null hypothesis, and the plot on the right is under the alternative.

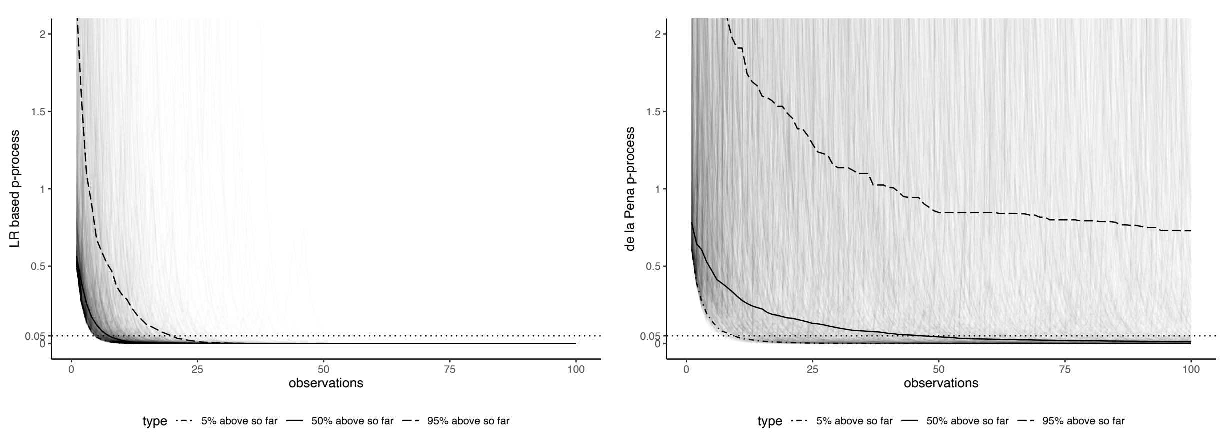

Figure 2: Plots of 1 000 post-hoc p-processes over the number of arrived observations under a normal alternative with mean $m = 1$. The highlighted lines are running quantiles: x% of the p-processes have not crossed below the line at the indicated time. The plot on the left is for our likelihood-ratio based p-process, and the plot on the right is for the one based on de la Peña [1999].

8 Acknowledgements

We would like to thank Sam van Meer, Muriel Pérez-Ortiz, Tyron Lardy, Will Hartog, Yaniv Romano, Jake Soloff, Stan Koobs, Peter Grünwald, Wouter Koolen and Aaditya Ramdas for useful discussions and comments. Part of this research was performed while the author was visiting the Institute for Mathematical and Statistical Innovation (IMSI), which is supported by the National Science Foundation (Grant No. DMS-1929348).

References

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

J. H. Chung and D. A. S. Fraser. Randomization tests for a multivariate two-sample problem. *Journal of the American Statistical Association*, 53(283):729–735, 1958. ISSN 01621459.

V. H. de la Peña. A general class of exponential inequalities for martingales and ratios. *The Annals of Probability*, 27(1):537–564, 1999.

M. L. Eaton. Group invariance applications in statistics. IMS, 1989.

P. Grünwald. Beyond neyman-pearson. *arXiv preprint arXiv:2205.00901*, 2023.

P. Grünwald, R. de Heide, and W. Koolen. Safe testing. *arXiv preprint arXiv:1906.07801*, 2023.

J. Hemerik and J. Goeman. Exact testing with random permutations. *Test*, 27(4):811–825, 2018.

S. R. Howard, A. Ramdas, J. McAuliffe, and J. Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055 – 1080, 2021. doi: 10.1214/20-AOS1991.

N. Ignatiadis, R. Wang, and A. Ramdas. E-values as unnormalized weights in multiple testing. *Biometrika*, page asad057, 09 2023. ISSN 1464-3510. doi: 10.1093/biomet/asad057.

O. Kallenberg. *Random measures, theory and applications*, volume 1. Springer, 2017.

N. W. Koning. More power by using fewer permutations. *arXiv preprint arXiv:2307.12832*, 2023.

N. W. Koning. Post-hoc p-values. *arXiv preprint arXiv:2312.08040*, 2024.

N. W. Koning and J. Hemerik. More Efficient Exact Group Invariance Testing: using a Representative Subgroup. *Biometrika*, 09 2023. ISSN 1464-3510. doi: 10.1093/biomet/asad050.

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

T. Lardy and M. F. Pérez-Ortiz. Anytime-valid tests of group invariance through conformal prediction. *arXiv preprint arXiv:2401.15461*, 2024.

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

M. F. Pérez-Ortiz, T. Lardy, R. de Heide, and P. Grünwald. E-statistics, group invariance and anytime valid testing. *arXiv preprint arXiv:2208.07610*, 2022.

A. Ramdas, J. Ruf, M. Larsson, and W. Koolen. Admissible anytime-valid sequential inference must rely on nonnegative martingales. *arXiv preprint arXiv:2009.03167*, 2022a.

A. Ramdas, J. Ruf, M. Larsson, and W. M. Koolen. Testing exchangeability: Fork-convexity, supermartingales and e-processes. *International Journal of Approximate Reasoning*, 141:83–109, 2022b.

A. Ramdas, P. Grünwald, V. Vovk, and G. Shafer. Game-theoretic statistics and safe anytime-valid inference. *arXiv preprint arXiv:2210.01948*, 2023.

G. Shafer. Testing by Betting: A Strategy for Statistical and Scientific Communication. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(2):407–431, 05 2021. ISSN 0964-1998. doi: 10.1111/rssa.12647.

G. Shafer and V. Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

V. Vovk. Testing randomness online. *Statistical Science*, 36(4):595–611, 2021.

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

V. Vovk. Testing exchangeability in the batch mode with e-values and markov alternatives. *arXiv preprint arXiv:2305.05284*, 2023b.

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

R. Wang and A. Ramdas. False discovery rate control with e-values. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(3):822–852, 2022.

A Conditional distributions

The goal of this section is to characterize the conditional distribution of Y given $h(Y)$ and its orbit O_Y. Let $\mathcal{Y}$ be our sample space on which our group $\mathcal{G}$ acts, and let h be some continuous function from $\mathcal{Y}$ into some Hausdorff space so that it induces a group action on this codomain $h(Gy) = Gh(y)$.

Let us consider the subset $\mathcal{K}^h$ of $\mathcal{G}$ that stabilizes a statistic h of the data:

$$\mathcal{K}^h(y) = \{G \in \mathcal{G} : h(Gy) = h(y)\}, \text{ where } y \in \mathcal{Y}.$$

Lemma 1 shows that this set is a compact subgroup of $\mathcal{G}$, henceforth referred to as its *h-stabilizing* subgroup. See Appendix D.8 for a proof.

Lemma 1. *$\mathcal{K}^h(y)$ is a compact subgroup of $\mathcal{G}$.*

This subgroup characterizes the conditional distribution of Y given an orbit O and $h(Y)$, as described in Proposition 3. In particular, given a draw $Y = y$, the conditional distribution of Y given $h(Y)$ and O_Y is equal to the distribution of $\overline{K}^h y$, where $\overline{K}^h$ is uniform (Haar) distributed in $\mathcal{K}^h(y)$. A proof is given in Appendix D.9.

Proposition 3. *Let $\overline{K}^h$ be uniform on $\mathcal{K}^h(Y)$. The conditional distribution of Y given $h(Y)$ and some orbit $O \in \mathcal{Y}/\mathcal{G}$, is equal to the distribution of $\overline{K}^h z$, where $z \in O$ and $h(Gz) = h(z)$ for some $G \in \mathcal{G}$.*

B Example: LR for exchangeability

In this section, we discuss a toy example of permutations on a small and finite sample space. While not as statistically interesting as the examples in Section 6, it is more tangible as the group itself is finite and easy to understand.

B.1 Exchangeability on a finite sample space

Suppose our sample space $\mathcal{Y}$ consists of the vectors $[1, 2, 3]$, $[1, 1, 2]$ and all their permutations. As a group $\mathcal{G}$, we consider the permutations on 3 elements, which we will denote by $\{abc, acb, bac, bca, cab, cba\}$. For example, *bac* represents the permutation that swaps the first two elements.

The orbits are then given by all permutations of $[1, 2, 3]$ and $[1, 1, 2]$

$$O_{[1,1,2]} = \{[1, 1, 2], [1, 2, 1], [2, 1, 1]\},$$

and

$$O_{[1,2,3]} = \{[1, 2, 3], [1, 3, 2], [2, 1, 3], [2, 3, 1], [3, 1, 2], [3, 2, 1]\}.$$

As $\mathcal{Y}$-valued orbit representatives, we pick the unique element in the orbit that is sorted in ascending order: $[1, 1, 2]$ and $[1, 2, 3]$. Notice that the group action of $\mathcal{G}$ is free on $O_{[1,2,3]}$ but not free on $O_{[1,1,2]}$.

For simplicity, let us restrict ourselves to $O_{[1,2,3]}$ first. On this orbit, the map γ is defined as the unique permutation that brings the element $[1, 2, 3]$ to $z \in O_{[1,2,3]}$. Moreover, on this orbit, the null hypothesis then states that $\gamma(Y)$ is uniform on the permutations, which in this case is equivalent to the hypothesis that Y is uniform on $O_{[1,2,3]}$.

Now let us restrict ourselves to $O_{[1,1,2]}$. On this orbit, there are multiple permutations that may bring a given element back to $[1, 1, 2]$. For example, both *bac*, as well as the identity permutation *abc* bring $[1, 1, 2]$ to itself. More generally, any permutation that brings $[1,1,2]$ to $z \in O_{[1,1,2]}$, can be preceded by bac, and the result still brings $[1,1,2]$ to $z \in O_{[1,1,2]}$. Even more abstractly speaking, let $\mathcal{S}_{[y]} = \{G \in \mathcal{G} : G[y] = [y]\}$ be the stabilizer subgroup of $[y]$ (the subgroup that leaves $[y]$ unchanged). Then, if $G^* \in \mathcal{G}$ carries $[y]$ to y, so does any element of $G^*\mathcal{S}_{[y]}$.

To construct γ on $O_{[1,1,2]}$, let $\overline{S}_{[y]}$ denote a uniform distribution on $\{abc, bac\}$, which is also well-defined in the general case as $\mathcal{S}_{[y]}$ is a compact subgroup and so admits a Haar probability measure. Moreover, let G_y be an arbitrary permutation that carries $[y]$ to y, say $G_{[1,1,2]} = abc$, $G_{[1,2,1]} = acb$ and $G_{[2,1,1]} = cba$. Then, we define $\gamma(y) = G_y\overline{S}_{[y]}$. Concretely, this means that $\gamma([1,1,2]) \sim \text{Unif}(abc, bac)$, $\gamma([1,2,1]) \sim \text{Unif}(acb, bca)$ and $\gamma([2,1,1]) \sim \text{Unif}(cba, cab)$. If Y is indeed uniform on $O_{[1,1,2]}$, then G_Y is uniform on $\{abc, acb, cba\}$ and so $\gamma(Y)$ is uniform on $\mathcal{G}$.

The definition of γ on the sample space $\mathcal{Y} = O_{[1,2,3]} \cup O_{[1,1,2]}$ is obtained by combining the definitions on the two separate orbits.

B.2 Likelihood ratios

We start with the orbit $O_{[1,2,3]}$. Suppose that our alternative distribution $P^1_{\mathcal{Y}}$ conditional on $O_{[1,2,3]}$ is that Y is uniform on $\{[1,2,3],[1,3,2]\}$ and all other arrangements happen with probability 0. As a density on $O_{[1,2,3]}$, we find

$$\begin{cases} 0.5, & \text{if } y \in \{[1,2,3],[1,3,2]\}, \\ 0, & \text{otherwise.} \end{cases}$$

As the density under the null is $1/6$ for each arrangement, the likelihood ratio is given by

$$\begin{cases} 3, & \text{if } y \in \{[1,2,3],[1,3,2]\}, \\ 0, & \text{otherwise.} \end{cases}$$

Since the group action is free on the orbit $O_{[1,2,3]}$, γ is a bijection between $O_{[1,2,3]}$ and the group, so likelihood ratio is

$$\begin{cases} 3, & \text{if } G \in \{abc, acb\}, \\ 0, & \text{otherwise.} \end{cases}$$

Now let us consider the orbit $O_{[1,1,2]}$. Suppose that our alternative $P^1_{\mathcal{Y}}$ conditional on $O_{[1,1,2]}$ is that Y equals $[1,1,2]$ with probability 1. The likelihood ratio on our orbit then becomes

$$\begin{cases} 3, & \text{if } y = [1,1,2], \\ 0, & \text{otherwise.} \end{cases}$$

In this case, the group action is not free, as both abc and bac are permutations that carry $[1,1,2]$ to itself. As a consequence $\gamma([1,1,2]) \sim \text{Unif}(abc, bac)$. This induces the following likelihood ratio on $\mathcal{G}$:

$$\begin{cases} 3, & \text{if } G \in \{abc, bac\}, \\ 0, & \text{otherwise.} \end{cases}$$

Now let us consider a likelihood ratio on $\mathcal{Y}$. For this, it is insufficient that we have an alternative on both $O_{[1,2,3]}$ and $O_{[1,1,2]}$, separately. We need to additionally specify the probability that that Y lands in $O_{[1,2,3]}$ and $O_{[1,1,2]}$ under the alternative. For simplicity, let us assume that the probability of each orbit is 1/2. The likelihood ratio on $\mathcal{Y}$ can then be derived to be

$$\begin{cases} 3/2, & \text{if } y \in \{[1,2,3],[3,2,1],[1,1,2]\}, \\ 0, & \text{otherwise.} \end{cases}$$

The likelihood ratio this induces on $\mathcal{G}$ is

$$\begin{cases} 3/2, & \text{if } G \in \{abc, bac\}, \\ 3, & \text{if } G = abc, \\ 0, & \text{otherwise,} \end{cases}$$

which is exactly the weighted average of the likelihood ratios on $\mathcal{G}$ that were induced on the individual orbits, weighted by the probability of each orbit.

C Finding the normalization constant analytically

For the likelihood ratio in Sections 6.2 and 6.3, we can easily compute the normalization constant under sphericity. The key trick is to use the fact that $\iota'\overline{G}y/\|y\|_2$ follows a Beta$(\frac{d-1}{2},\frac{d-1}{2})$ distribution on the interval $[-1,1]$. The normalization constant is then equal to its moment generating function:

$$\mathbb{E}_{\overline{G}}\exp(\mu\iota'\overline{G}y) = 1 + \mu\|y\|_2\mathbb{E}_{\overline{G}}\widetilde{B} + \mu^2\|y\|_2^2\mathbb{E}_{\overline{G}}\widetilde{B}^2/2! + \dots,$$

where $\widetilde{B} \sim$ Beta$\left(\frac{d-1}{2},\frac{d-1}{2}\right)$ on $[-1,1]$. For this generalized beta distribution, the odd moments are all 0, since it is symmetric about 0. Moreover, the even moments are given by

$$\mathbb{E}_{\overline{G}}(\widetilde{B})^{k+2} = \mathbb{E}_{\overline{G}}(\widetilde{B})^k(k-1)/(n+k-2),$$

with $\mathbb{E}_{\overline{G}}(\widetilde{B})^0 = 1$ and n the dimension of y. This means the normalization constant can be easily approximated to high precision, which we exploit in our simulation studies.

Furthermore, it is possible to numerically stabilize the computations by using the fact that

$$1/n\sum_i^n Z_i = \max_i \log(Z_i) + \log(1/n\sum_{i=1}^n(\exp(\log(Z_i) - \max_i \log(Z_i)))).$$

D Omitted proofs

D.1 Proof of Theorem 1

Proof. The proof strategy is to show the p-value is exact on every orbit, which then implies that it is exact for any mixture over orbits as well.

Let $O \in \mathcal{Y}/\mathcal{G}$ be some arbitrary orbit. Let Z be a random variable that is uniform on O. Let z be some arbitrary element in O. First, observe that $\overline{G}Z \overset{d}{=} \overline{G}z$ is uniform on O regardless of the distribution of Z. As a consequence, we have

$$\begin{aligned} p(Z) &= \mathbb{P}_{\overline{G}}\left(T(\overline{G}Z) > T(Z)\right) + \overline{u}\mathbb{P}_{\overline{G}}\left(T(\overline{G}Z) = T(Z)\right) \\ &= \mathbb{P}_{\overline{G}}\left(T(\overline{G}z) > T(Z)\right) + \overline{u}\mathbb{P}_{\overline{G}}\left(T(\overline{G}z) = T(Z)\right) \end{aligned}$$

Then, as Z is $\mathcal{G}$ invariant through T, we have $T(Z) \overset{d}{=} T(\overline{G}^* Z)$, where $\overline{G}^*$ is uniform on $\mathcal{G}$. This implies

$$\begin{aligned} \mathbb{P}_{\overline{G}}\left(T(\overline{G}z) > T(Z)\right) + \overline{u}\mathbb{P}_{\overline{G}}\left(T(\overline{G}z) = T(Z)\right) &\overset{d}{=} \mathbb{P}_{\overline{G}}\left(T(\overline{G}z) > T(\overline{G}^* Z)\right) + \overline{u}\mathbb{P}_{\overline{G}}\left(T(\overline{G}z) = T(\overline{G}^* Z)\right) \\ &\overset{d}{=} \mathbb{P}_{\overline{G}}\left(T(\overline{G}z) > T(\overline{G}^* z)\right) + \overline{u}\mathbb{P}_{\overline{G}}\left(T(\overline{G}z) = T(\overline{G}^* z)\right), \end{aligned}$$

where the second equality again follows from $\overline{G}^* Z \overset{d}{=} \overline{G}^* z$.

Then, it is straightforward to show that $P_B(A > B) + \overline{u}P_B(A = B)$ is uniform on $[0, 1]$ for i.i.d. random variables A and B. As $\overline{G}$ and $\overline{G}^*$ are independent and identically distributed, the same holds for $T(\overline{G}z)$ and $T(\overline{G}^* z)$. This implies that $p(Z)$ is uniform on $[0, 1]$. As this holds for an arbitrary orbit O, it holds for every orbit. Then, as any mixture of random variables that are uniform on $[0, 1]$ is also uniform on $[0, 1]$, $p(Y)$ is uniform on $[0, 1]$. □

D.2 Proof of Theorem 2

Proof. We show that it is exact on every orbit, which implies that it is also exact if we mix over the orbits. Let $O \in \mathcal{Y}/\mathcal{G}$ be some arbitrary orbit. Let Z be a random variable on O that is uniform on O.

First, observe that $\overline{G}G \overset{d}{=} \overline{G}$ for all $G \in \mathcal{G}$, as $\overline{G}$ is a $\mathcal{G}$ invariant random variable. As a consequence, the map $z \mapsto \mathbb{E}_{\overline{G}}T(\overline{G}z)$ is $\mathcal{G}$ invariant: $\mathbb{E}_{\overline{G}}T(\overline{G}z) = \mathbb{E}_{\overline{G}}T(\overline{G}Gz)$ for all $G \in \mathcal{G}$. This implies that $\mathbb{E}_{\overline{G}}T(\overline{G}z)$ is constant on O. As Z only takes value on O, this means $\mathbb{E}_{\overline{G}}T(\overline{G}Z) = \mathbb{E}_{\overline{G}}T(\overline{G}z)$. As a result,

$$\mathbb{E}_Z\left[T(Z)/\mathbb{E}_{\overline{G}}T(\overline{G}Z)\right] = \mathbb{E}_Z T(Z)/\mathbb{E}_{\overline{G}}T(\overline{G}z),$$

Now, as Y is $\mathcal{G}$ invariant through T, we have that $T(Z) \overset{d}{=} T(\overline{G_2}Z)$, where $\overline{G}_2$ is uniform on $\mathcal{G}$ independently from Z. As a consequence,

$$\mathbb{E}_Z T(Z)/\mathbb{E}_{\overline{G}}T(\overline{G}z) = \mathbb{E}_{\overline{G}_2 Z}T(\overline{G}_2 Z)/\mathbb{E}_{\overline{G}}T(\overline{G}z).$$

As $\overline{G}_2$ and Z are independent, Tonelli's theorem gives

$$\mathbb{E}_{\overline{G}_2 Z}T(\overline{G}_2 Z)/\mathbb{E}_{\overline{G}}T(\overline{G}z) = \mathbb{E}_Z\mathbb{E}_{\overline{G}_2}T(\overline{G}_2 Z)/\mathbb{E}_{\overline{G}}T(\overline{G}z).$$

Then, we can again use the $\mathcal{G}$ invariance of the map $z \mapsto \mathbb{E}_{\overline{G}_2}T(\overline{G}_2 z)$, to argue that $\mathbb{E}_{\overline{G}_2}T(\overline{G}_2 Z) = \mathbb{E}_{\overline{G}_2}T(\overline{G}_2 z)$. As a consequence,

$$\mathbb{E}_Z\mathbb{E}_{\overline{G}_2}T(\overline{G}_2 Z)/\mathbb{E}_{\overline{G}}T(\overline{G}Z) = \mathbb{E}_{\overline{G}_2}T(\overline{G}_2 z)/\mathbb{E}_{\overline{G}}T(\overline{G}z) = 1.$$

As O was arbitrarily chosen, this holds for every orbit in $\mathcal{X}/\mathcal{G}$. In turn, this implies that it also holds unconditionally.

Next, we show that every exact post-hoc p-value is of the form $\mathfrak{p}_T$. Suppose that $\mathfrak{p}(Y)$ is an exact post-hoc p-value for $\mathcal{G}$ invariance of Y through $\mathfrak{p}$. This means that it is exact for every random variable that is $\mathcal{G}$ invariant through $\mathfrak{p}$.

Let us pick an arbitrary orbit O, and take a random variable Z that is uniformly distributed on this orbit. This random variable is clearly $\mathcal{G}$ invariant. As a consequence, $\mathfrak{p}(Z)$ is an exact post-hoc p-value through $\mathfrak{p}$:

$$1 = \mathbb{E}_Z[1/\mathfrak{p}(Z)] = \mathbb{E}_Z\mathbb{E}_{\overline{G}}[1/\mathfrak{p}(\overline{G}Z)].$$

Now, notice that the map $z \mapsto \mathbb{E}_{\overline{G}}[1/\mathfrak{p}(\overline{G}z)]$ is $\mathcal{G}$ invariant. As a consequence, it is constant on the orbit O. As Z only takes value on O, this implies $\mathbb{E}_{\overline{G}}[1/\mathfrak{p}(\overline{G}Z)]$ is the same for any draw of Z. This implies $\mathbb{E}_Z\mathbb{E}_{\overline{G}}[1/\mathfrak{p}(\overline{G}Z)] = \mathbb{E}_{\overline{G}}[1/\mathfrak{p}(\overline{G}Z)]$, which in turn implies that $\mathbb{E}_{\overline{G}}[1/\mathfrak{p}(\overline{G}Z)] = 1$.

As the orbit was arbitrarily given, this holds for any orbit. As a consequence, $\mathbb{E}_{\overline{G}}[1/\mathfrak{p}(\overline{G}z)] = 1$ regardless of the orbit that z is on. This implies $\mathbb{E}_{\overline{G}}[1/\mathfrak{p}(\overline{G}Y)] = 1$.

Finally, we have that $\mathfrak{p}(Y)$ is indeed of the conjectured form as

$$\mathfrak{p}(Y) = \mathfrak{p}(Y)/\mathbb{E}_{\overline{G}}[1/\mathfrak{p}(\overline{G}Y)] = \mathfrak{p}_{1/\mathfrak{p}}(Y).$$

Hence, if $\mathfrak{p}$ is an exact post-hoc p-value for $\mathcal{G}$ invariance through $\mathfrak{p}$, then it must be equal to some test statistic of the form $\mathfrak{p}_T$. □

D.3 Proof of Theorem 3

Proof. Conditionally on X^{n-1}, we have that X^n is $\mathcal{K}_n(X^{n-1})$ invariant through T_n: $T_n(X^n) \overset{d}{=} T_n(\overline{K}_n X^{n-1})$ conditional on $X^n \in O_{X^n}$, where $\overline{K}_n$ is uniform on $\mathcal{K}_n(X^n)$. As a consequence, by Theorem 2, we have

$$\mathbb{E}_{T_n(X^n)}\left(\frac{T_n(X^n)}{\mathbb{E}_{\overline{K}_n} T_n(\overline{K}_n X^n)} \;\Big|\; X^{n-1}\right) = 1.$$

This means that $\mathfrak{p}_n(X^n)$ is an exact post-hoc p-value conditional on X^{n-1}, $n \geq 2$. Moreover, $\mathfrak{p}_1(X^1)$ is a 'plain' exact post-hoc p-value as $\mathcal{K}_1 = \mathcal{G}_1$. Applying the law of iterated expectations, we have

$$
\begin{aligned}
\mathbb{E}1/\mathfrak{p}_n(X^n) &= \mathbb{E}\left[\prod_{i=1}^{n} \frac{T_i(X^i)}{\mathbb{E}_{\overline{K}_i} T_i(\overline{K}_i X^i)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\frac{T_n(X^n)}{\mathbb{E}_{\overline{K}_n} T_n(\overline{K}_n X^n)} \prod_{i=1}^{n-1} \frac{T_i(X^i)}{\mathbb{E}_{\overline{K}_i} T_i(\overline{K}_i X^i)} \;\Big|\; X^{n-1}\right)\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{n-1} \frac{T_i(X^i)}{\mathbb{E}_{\overline{K}_i} T_i(\overline{K}_i X^i)} \mathbb{E}\left(\frac{T_n(X^n)}{\mathbb{E}_{\overline{K}_n} T_n(\overline{K}_n X^n)} \;\Big|\; X^{n-1}\right)\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{n-1} \frac{T_i(X^i)}{\mathbb{E}_{\overline{K}_i} T_i(\overline{K}_i X^i)}\right] \\
&= \mathbb{E}\left[\frac{T_1(X^1)}{\mathbb{E}_{\overline{G}_1} T_i(\overline{G}_1 X^1)}\right] = 1,
\end{aligned}
$$

where the second-to-last equality follows from induction. □

D.4 Proof of Theorem 4

Proof. The proof is similar to that of Theorem 3.

Conditionally on $S_{n-1}(X^{n-1})$, we have that X^n is $\mathcal{F}_n$ invariant through S_n: $S_n(X^n) \overset{d}{=} S_n(\overline{F}_n^* X^n)$ conditional on $S_{n-1}(X^{n-1})$ and $O_{X^n}^{\mathcal{F}_n}$, where $\overline{F}_n^*$ is uniform on $\mathcal{F}_n$. As a consequence, Theorem 2 implies

$$
\mathbb{E}_{S_n(X^n)}\left[\frac{T_n(S_n(X^n))}{\mathbb{E}_{\overline{F}_n} T_n(S_n(\overline{F}_n X^n))} \;\Big|\; S_{n-1}(X^{n-1})\right] = 1.
$$

This means $\mathfrak{p}_n(X^n)$ is an exact post-hoc p-value conditional on $S_{n-1}(X^{n-1})$, $n \geq 2$. Moreover, $\mathfrak{p}_1(X^1)$ is an exact post-hoc p-value, unconditionally. We can then apply the iterated expectations strategy from the final part of the proof of Theorem 3 in order to conclude that $(\mathfrak{p}_n)_{n\geq 1}$ is an exact post-hoc p-process. □

D.5 Proof of Proposition 2

Proof. As $\mathcal{F}_{n-1}$ is a subgroup of $\mathcal{F}_n$, its action on $\mathcal{X}^n$ is well-defined. The $\mathcal{F}_{n-1}$ invariance of S_n means that $S_n(Fx^n) = S_n(x^n)$ for all $F \in \mathcal{F}_{n-1}$ and $x^n \in \mathcal{X}^n$. As a result, S_n is constant on any orbit in $\mathcal{X}^n/\mathcal{F}_{n-1}$.

Next, $x^{n-1} \in \mathcal{X}^n$, as $\mathcal{X}^{n-1} \subseteq \mathcal{X}^n$. Hence, the orbit of x^{n-1} under $\mathcal{F}_{n-1}$ is in $\mathcal{X}^n/\mathcal{F}_{n-1}$. As a consequence, S_n is constant on the orbit of x^{n-1}. In turn, this means that $S_n(X^n)$ is independent of X^{n-1}. □

D.6 Proof of Theorem 5

Proof of Theorem 5. By Lemma 2, $S_n(X^n)$ is independent of X^{n-1}. This implies $S_n(X^n)$ is also independent of $S_{n-1}(X^{n-1})$. Moreover, $S_n(X^n)$ has the same distribution on every orbit in $\mathcal{X}^n/\mathcal{F}_n$. As a consequence, it is independent of O_{X^n}. Hence, the conditional distribution of $S_n(X^n)$ given $\sigma(S_{n-1}(X^{n-1}), O_{X^n})$ is equal to its unconditional distribution.

Let $x^n \in \mathcal{X}^n$ be some arbitrary point. Then, $\overline{F}_n x^n$ is uniform on the orbit of x^n. As $S_n(X^n)$ has the same distribution on every orbit and X^n is $\mathcal{F}_n$ invariant through S_n, this distribution equals that of $S_n(\overline{F}_n x^n)$. Finally, as X^n itself is a point in $\mathcal{X}^n$, $S_n(X^n) \stackrel{d}{=} S_n(\overline{F}X^n)$. □

D.7 Proof of Theorem 6

Proof. For notational convenience, let us define a function f as $f(y) = \mathbb{E}_{\overline{G}} d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}y)$. Observe that f is a $\mathcal{G}$ invariant function as

$$f(Gy) = \mathbb{E}_{\overline{G}} d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}y) = \mathbb{E}_{\overline{G}} d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}Gy) = f(y),$$

since $\overline{G} \stackrel{d}{=} \overline{G}G$ as $\overline{G}$ is a $\mathcal{G}$ invariant random variable. As a consequence, f is constant on the orbit $\{z \in \mathcal{Y} : y = Gz, \text{ for some } G\}$ of y. As $0 < f(y) < \infty$, the function f is proportional to a uniform distribution on every orbit. Now, the statistic integrates to 1 on every orbit as

$$\mathbb{E}_{\overline{G}_2} \frac{d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}_2 y)}{\mathbb{E}_{\overline{G}} d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}\overline{G}_2 y)} = \mathbb{E}_{\overline{G}_2} \frac{d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}_2 y)}{\mathbb{E}_{\overline{G}} d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}y)} = \frac{\mathbb{E}_{\overline{G}_2} d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}_2 y)}{\mathbb{E}_{\overline{G}} d\mathbb{P}_{\mathcal{Y}}/d\lambda(\overline{G}y)} = 1,$$

where $\overline{G}_2$ is uniform on $\mathcal{G}$, independently from $\overline{G}$. Hence, for every orbit, the statistic is a likelihood ratio statistic for testing a uniform distribution on an orbit against $\mathbb{P}_{\mathcal{Y}}$. Finally, as $\mathcal{G}$ invariance is equivalent to being uniform on every orbit, the statistic is a likelihood ratio statistic for testing $\mathbb{P}_{\mathcal{Y}}$ against $\mathcal{G}$ invariance. □

D.8 Proof of Lemma 1

Proof. We start with showing that $\mathcal{K}^h(y)$ is a subgroup. This follows from the fact that it a stabilizer subgroup of $\mathcal{G}$ that stabilizes $h(y)$, but we prove it for completeness. Then, we show it is closed, so that by the compactness of $\mathcal{G}$ it is also compact.

First observe that the identity element is in $\mathcal{K}^h(y)$. Suppose that $K_1, K_2 \in \mathcal{K}^h(y)$. Then,

$$K_1 K_2 h(y) = K_1 h(y) = h(y),$$

so that $\mathcal{K}^h(y)$ is closed under compositions. Moreover, it is closed under inverses as for any $K \in \mathcal{K}^h(y)$

$$h(y) = Ih(y) = K^{-1}Kh(y) = K^{-1}h(y) = h(K^{-1}y),$$

so that $K^{-1} \in \mathcal{K}^h(y)$. This means that $\mathcal{K}^h(y)$ is closed under inverses. Hence, $\mathcal{K}^h(y)$ is a subgroup of $\mathcal{G}$.

Finally, we show that $\mathcal{K}^h(y)$ is topologically closed. Define the map $f_y : \mathcal{G} \to \mathcal{Y}$ as the composition between h and the group action: $f_y(G) = h(Gy)$. As both h and the group action are continuous, the composition f_y is also continuous. Since the space in which $\{h(y)\}$ lives is Hausdorff, it is also a T_1 space so that $\{h(y)\}$ is closed. Hence, $\mathcal{K}^h(y)$ is the pre-image of the closed set $\{h(y)\}$ under a continuous map, and so $\mathcal{K}^h(y)$ is also closed, and therefore compact. □

D.9 Proof of Proposition 3

Proof. Let us start by fixing an arbitrary orbit O under $\mathcal{G}$ acting on $\mathcal{Y}$.

As Y is $\mathcal{G}$ invariant, it is also invariant under any of its compact subgroups. By Lemma 1, $\mathcal{S}$ is such a compact subgroup. As a consequence, Y is uniform on the (sub)orbits of $\mathcal{S}$ acting on O.

Let us condition on the suborbit in which Y falls. This suborbit consists of all the points y for which $h(y) = h(GY)$ for some $G \in \mathcal{G}$. As $\mathcal{G}$ acts transitively on O, any point on the suborbit can be reached by applying a transformation G to Y. Hence, the suborbit of Y under $\mathcal{S}$ consists exactly of those elements $y \in O$ for which $h(y) = h(Y)$. This is exactly the subset to which we restrict Y by conditioning on $h(Y)$. As a result, the conditional distribution of Y given O and $h(Y)$ is uniform on this suborbit.

This distribution can be characterized as the distribution of $\overline{S}z$, where $\overline{S}$ is uniform on $\mathcal{S}$, and z is some element on the suborbit. □