Supplement to "Statistical Disclosure Control for External AI via Transformation Groups"

Lingxuan Shao

This supplementary material to "Statistical Disclosure Control for External AI via Transformation Groups" collects the simulations of the hypothesis testing framework, the necessary notations and mathematical preliminaries, the dataset and prompts used in the example, and the proofs of the lemmas and theorems.

S1 Simulation Studies

In this section, we conduct simulation experiments to empirically validate the hypothesis testing framework for local traceability developed in Section 5. While in practice the external AI operator A is treated as a black box, our simulations require specifying concrete statistical methods to explicitly control whether the null hypothesis H_0 (local traceability holds) or the alternative H_1 (local traceability fails) is true. The purpose is therefore a size-and-power study of the test itself and not an assessment of the applicability of the framework to any particular AI service; the latter is the object of the pottery example in Section 7 of the main text.

S1.1 The distance-based permutation test

Throughout the simulation study and the pottery-company example of the main text, the multi-sample test of Section 5 is instantiated as a distance-based permutation test with the pseudo-F statistic (Arboretti et al. 2025). We describe the statistic and the p-value explicitly here.

Let $Y_{ij} = R_{T_i}\big(A_{ij}\{T_i(X_0)\}\big) \in \mathbb{R}^q$, $i = 1, \dots, n$, $j = 1, \dots, m$, be the recovered outputs and write $N = nm$. Because the coordinates of Y_{ij} may have different scales, each coordinate is standardized before distances are computed:

$$\widetilde{Y}_{ij,k} = \frac{Y_{ij,k}}{s_k}, \qquad s_k^2 = \frac{1}{N-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\big(Y_{ij,k} - \bar{Y}_{\cdot,k}\big)^2, \qquad k = 1, \dots, q, \tag{S1.1}$$

where $\bar{Y}_{\cdot,k}$ is the average of the kth coordinate over all $N = nm$ recovered outputs; coordinates with $s_k = 0$ are left unscaled. The squared Euclidean distance between two recovered outputs is then

$$d^2(a, b) = \sum_{k=1}^{q}\big(\widetilde{Y}_{a,k} - \widetilde{Y}_{b,k}\big)^2. \tag{S1.2}$$

Writing $I_i = \{im - m + 1, \dots, im\}$ for the indices belonging to the ith transformation, the total and within-group sums of squares and the pseudo-F statistic are

$$\begin{gathered}\mathrm{SST} = \frac{1}{2N}\sum_{a=1}^{N}\sum_{b=1}^{N} d^2(a, b), \qquad \mathrm{SSW} = \frac{1}{2m}\sum_{i=1}^{n}\sum_{a\in I_i}\sum_{b\in I_i} d^2(a, b), \\ F = \frac{(\mathrm{SST} - \mathrm{SSW})/(n-1)}{\mathrm{SSW}/(N-n)}.\end{gathered} \tag{S1.3}$$

The permutation distribution of F is obtained by randomly reassigning the N observations to n groups of sizes $m, \dots, m$, which is the distribution of F under the null hypothesis of

local traceability. With B permutations and the observed statistic F_{obs}, the p-value is

$$\widehat{p} = \frac{1 + \sum_{b=1}^{B} \mathbf{1}\{F_b^* \geq F_{\text{obs}}\}}{B+1}, \tag{S1.4}$$

where F_b^* denotes the pseudo-F statistic computed on the bth permutation. We use $B = 999$ throughout, so that the smallest attainable p-value is 0.001.

S1.2 Simulation settings and power

We consider two distinct settings, each corresponding to a prototypical transformation group and a realistic prediction task.

- **Setting one: One-dimensional positive time-series prediction with the scaling group.** The task is to forecast the next value of a positive univariate time series. We fix a historical dataset of length $K = 10$:

  $$X_0 = (x_1, \ldots, x_K) = (1.2,\ 1.5,\ 1.8,\ 1.3,\ 1.7,\ 1.6,\ 1.9,\ 2.1,\ 1.8,\ 2.0).$$

  The transformation group is the scaling group $G_{\text{scale}} \cong (\mathbb{R}_{>0}, \times)$, with sampling distribution as specified in Example 3.

  - Under H_0, the AI operator is designed to be homogeneous of degree one. We adopt an ARIMA(1,1,0) model (Montgomery et al. 2024): the differenced series $\Delta x_k = x_k - x_{k-1}$ for $k = 2, \ldots, K$ is modeled by an AR(1) process $\Delta x_k = \phi \Delta x_{k-1} + \varepsilon_k$. The coefficient ϕ is estimated via ordinary least squares, and the stochastic forecast is generated by the one-step prediction plus a random residual drawn from a normal distribution with

variance estimated from the empirical residuals. The output is thresholded to ensure positivity.

- Under H_1, the scale homogeneity is broken by the intended AI model bias. The forecast is constructed by taking a convex combination between the forecast produced under the H_0 method and a fixed constant $\xi = 1$: $\text{Forecast} = (1-\eta) \cdot \text{Forecast}_{H_0} + \eta \cdot \xi$, where ξ may result from the political bias or inserted advertisement, and $\eta \in (0, 1)$ is a mixing proportion.

- **Setting two: Two-dimensional location prediction with the rotation group.** The task is to predict the next vector in a bivariate time series. We fix a historical dataset of length $K = 10$:

$$X_0 = (\mathbf{x}_1, \ldots, \mathbf{x}_K)^\top = \begin{pmatrix} 0.5 & 0.7 & 1.0 & 1.3 & 1.7 & 2.0 & 2.4 & 2.9 & 3.3 & 3.8 \\ 0.3 & 0.5 & 0.8 & 1.2 & 1.5 & 1.8 & 2.2 & 2.6 & 3.0 & 3.5 \end{pmatrix}^\top$$

The transformation group is the rotation group $G_{\text{SO}} \cong \text{SO}(2)$, with the uniform (Haar) distribution on angles $\theta \in [0, 2\pi)$.

- Under H_0, the AI operator is designed to be rotation-equivariant. We fit a restricted VAR(1) model $\mathbf{x}_k = \phi I_2 \mathbf{x}_{k-1} + \boldsymbol{\epsilon}_k$ via constrained ordinary least squares, yielding an estimate $\widehat{\phi}$, and the stochastic forecast is generated by the one-step prediction plus a random residual drawn from a multivariate normal distribution with covariance matrix estimated from the empirical residuals.

- Under H_1, rotation equivariance is broken using a similar convex-combination pertur-

bation as in Setting one. The forecast is computed as $(1-\eta)\cdot \text{Forecast}_{H_0} + \eta \cdot \boldsymbol{\xi}$, where $\boldsymbol{\xi} = (1,1)^\top$ may result from the political bias or inserted advertisement, and $\eta \in (0,1)$ is a mixing proportion.

The simulation proceeds according to the following common steps for both settings:

- Fix the historical dataset X_0 as specified above.
- Sample n independent transformations from the group's probability measure μ.
- For each sampled transformation T_i, generate m independent realizations of the AI output $A_{ij}\{T_i(X_0)\}$ using the procedure described under the chosen hypothesis (H_0 or H_1).
- Apply the corresponding recovery map R_{T_i} to obtain the processed data $Y_{ij} = R_{T_i}(A_{ij}\{T_i(X_0)\})$.
- Perform a distance-based permutation test with pseudo-F statistic (Arboretti et al. 2025) on the n groups $\{Y_{ij}\}_{j=1}^m$ for $i = 1, \ldots, n$ at significance level $\alpha = 0.05$.

We run experiments for different sample-size combinations: $n = 5, 10$ and $m = 10, 20$, and for various mixing proportions η. The results presented in Table 1 demonstrate the empirical performance of the proposed hypothesis testing framework across both simulation settings at nominal level $\alpha = 0.05$. For each setting, we report the empirical size (under H_0, representing local traceability) and power (under H_1, representing violations of local traceability) at various mixing proportions η.

Table 1: Empirical size (under H_0) and power (under H_1) for both settings, based on 1000 replications.

Setting 1	n	m	H_0	H_1 ($\eta = 0.10$)	H_1 ($\eta = 0.25$)	H_1 ($\eta = 0.50$)
	5	10	0.051	0.310	0.826	0.993
	5	20	0.054	0.509	0.926	1.000
	10	10	0.044	0.492	0.975	1.000
	10	20	0.043	0.726	0.993	1.000
Setting 2	n	m	H_0	H_1 ($\eta = 0.05$)	H_1 ($\eta = 0.10$)	H_1 ($\eta = 0.20$)
	5	10	0.043	0.671	0.998	1.000
	5	20	0.036	0.955	0.997	1.000
	10	10	0.040	0.940	1.000	1.000
	10	20	0.037	0.999	1.000	1.000

The results confirm robust size control across both settings, with empirical size consistently close to $\alpha = 0.05$. The test exhibits high and systematically increasing power with larger sample sizes (n, m) and perturbation strength (η). Notably, the two-dimensional rotation group setting demonstrates substantially higher sensitivity than the one-dimensional scaling group, achieving near-perfect power even at minimal perturbations ($\eta = 0.05$), a benefit attributable to its richer geometric structure. Furthermore, increasing the number of repeated queries per transformation (m) provides a pronounced boost in power, particularly for smaller values of n.

Read the other way round, the same table gives the smallest deviation that the test can detect at each sample size, which is the operational meaning of admissibility (Definition 5 of

the main text). At the pottery configuration $(n, m) = (10, 20)$, a deviation of size $\eta = 0.25$ is detected with probability 0.993 and one of size $\eta = 0.10$ with probability 0.726; smaller deviations go essentially undetected. An admissibility claim should therefore be read at that resolution and quoted together with the transformation group and the design (n, m), since the resolution depends on the setting: in Setting 2, $\eta = 0.05$ is detected with probability 0.671 already at $(5, 10)$.

S2 Notations and Preliminaries

This section establishes the mathematical notation and reviews essential background concepts from group theory, topological groups, and Haar measures that underpin our framework.

S2.1 Mathematical Notation

Let $\mathcal{X}$ denote the *data space*. The true dataset of interest is denoted by $X_0 \in \mathcal{X}$. The stochastic AI operator is $A : \Omega \times \mathcal{X} \to \Theta$, where Ω is the underlying sample space and Θ is the *decision space*. For a fixed input X, we write $A\{X\}$ for the random output.

A *transformation* is a self-homeomorphism $t : \mathcal{X} \to \mathcal{X}$, and the collection of all such maps forms the group $\mathrm{Homeo}(\mathcal{X})$ under composition. Our framework employs a subgroup $G \subset \mathrm{Homeo}(\mathcal{X})$ from which transformations are sampled. For any $t \in G$, we denote by $R_t \in \mathrm{Homeo}(\Theta)$ a corresponding *recovery map* that recovers the original target statistic in

distribution. The notation $\stackrel{d}{=}$ denotes equality in distribution.

The *disclosure space* associated with a group G and dataset X is denoted by $\Theta_{GX} := \{\mathbb{E}[A\{t^{-1}(X)\}] \mid t \in G\}$, where the expectation is taken over the randomness of A. The symbol μ denotes a probability measure on G obtained from the Haar measure $\tilde{\mu}$ by reweighting with a positive weight function ω, as in Section 4 of the main text.

S2.2 Group Theory Essentials

A *group* $(G, \circ)$ is a set G equipped with a binary operation $\circ : G \times G \to G$ satisfying:

- *Associativity*: $(g \circ h) \circ k = g \circ (h \circ k)$ for all $g, h, k \in G$.
- *Identity*: There exists $e \in G$ such that $e \circ g = g \circ e = g$ for all $g \in G$.
- *Inverse*: For each $g \in G$, there exists $g^{-1} \in G$ with $g \circ g^{-1} = g^{-1} \circ g = e$.

A subset $H \subset G$ that is itself a group under the same operation is called a *subgroup*. A group G *acts* on a space $\mathcal{X}$ if there is a map $\varphi : G \times \mathcal{X} \to \mathcal{X}$, written as $(g, x) \mapsto g \cdot x$, satisfying $e \cdot x = x$ and $(g \circ h) \cdot x = g \cdot (h \cdot x)$. The *orbit* of a point $x \in \mathcal{X}$ under G is $G \cdot x = \{g \cdot x \mid g \in G\}$. The action is *transitive* if for any $x, y \in \mathcal{X}$ there exists $g \in G$ with $g \cdot x = y$ (i.e., the whole space is a single orbit). Our notion of local transitivity adapts this concept to the disclosure-space setting.

S2.3 Topological Groups

A *topological group* is a group G endowed with a topology such that the group operations $(g, h) \mapsto g \circ h$ and $g \mapsto g^{-1}$ are continuous. In our framework, the transformation groups are considered as subgroups of $\text{Homeo}(\mathcal{X})$ equipped with the *compact-open topology*, which is the topology generated by subbasic sets

$$V(K, U) = \{t \in \text{Homeo}(\mathcal{X}) \mid t(K) \subseteq U\},$$

where $K \subset \mathcal{X}$ is compact and $U \subset \mathcal{X}$ is open.

A topological space is *locally compact* if every point has a compact neighborhood. This property is essential for the existence of a Haar measure. A space is *Hausdorff* if distinct points have disjoint neighborhoods.

Some standard examples illustrate these notions. (i) $\mathbb{R}^p$ with the Euclidean topology is locally compact and Hausdorff: every point has a compact closed ball as a neighborhood, and distinct points are separated by disjoint open balls. (ii) The general linear group $\text{GL}(p, \mathbb{R})$ and the special orthogonal group $\text{SO}(p)$, regarded as subsets of $\mathbb{R}^{p^2}$ with the subspace topology, are topological groups, because matrix multiplication and inversion are continuous; $\text{SO}(p)$ is compact, whereas $\text{GL}(p, \mathbb{R})$ is locally compact but not compact. (iii) If $\mathcal{X}$ is locally compact, Hausdorff and locally connected, then $\text{Homeo}(\mathcal{X})$ equipped with the compact-open topology is a topological group; the subbasic neighborhoods $V(K, U)$ are those sets of homeomorphisms that map a compact set K into an open set U, so two homeomorphisms are

close when they agree on compact sets.

S2.4 Haar Measures on Locally Compact Groups

The cornerstone of our random sampling strategy is the Haar measure, which provides a uniform distribution over a topological group. Let G be a locally compact Hausdorff topological group. A *left Haar measure* $\tilde{\mu}$ is a non-zero Radon measure on G that is *left-invariant*: for every measurable set $E \subseteq G$ and every $g \in G$, $\tilde{\mu}(gE) = \tilde{\mu}(E)$, where $gE = \{g \circ h \mid h \in E\}$.

S3 Data set and prompts

The two prompts used in the experiments differ only in the description of the target quantity: Prompt 1 asks for the point forecast, and Prompt 2 for the 20% quantile, which the company reads as a floor that sales are expected to exceed with probability 0.8. In each of the four combinations of task and sampling measure considered in the main text, the placeholder "transformed data" and the 2026 features are replaced by their transformed versions.

Prompt 1. *You are given a dataset of a handicraft company's pottery collection sales from 2016 to 2025. Each year in the dataset includes the following variables: Year index; Average retail price per piece (USD); Annual marketing expenditure (thousand USD); Number of direct competitor artisan brands; Average customer rating on e-commerce platforms (1-5 scale); Pottery size (inches, diameter); Design coordinate x (relative units); Design coordinate y (relative units); Annual sales volume (pieces).*

Table 2: Dataset $X_0 = \{(X_i, Y_i)\}_{i=1}^{10}$ spanning 2016–2025 for the ceramic collection. Variables: $X_{i,[1]}$ = year index; $X_{i,[2]}$ = average retail price per piece (USD); $X_{i,[3]}$ = annual marketing expenditure (thousand USD); $X_{i,[4]}$ = number of direct competitor artisan brands; $X_{i,[5]}$ = average customer rating on e-commerce platforms (1–5 scale); $X_{i,[6]}$ = pottery size (inches, diameter); $X_{i,[7]}$ = horizontal design coordinate x (relative units); $X_{i,[8]}$ = vertical design coordinate y (relative units); Y_i = annual sales volume (pieces). The design coordinates (x, y) represent the relative position of a signature decorative element on each pottery piece.

*2021 price increase due to rising clay and glaze costs.

**2021 sales decline due to supply chain disruptions.

$X_{i,[1]}$	$X_{i,[2]}$ (\$)	$X_{i,[3]}$ (K\$)	$X_{i,[4]}$	$X_{i,[5]}$	$X_{i,[6]}$ (in)	$X_{i,[7]}$ (x)	$X_{i,[8]}$ (y)	Y_i (pieces)
Year	Price	Marketing	Competitors	Rating	Size	x_coord	y_coord	Sales
2016	299	45	12	4.1	5.2	0.8	0.2	2831
2017	289	52	15	4.0	5.5	0.7	0.3	3243
2018	279	58	18	3.9	5.7	0.6	0.4	3528
2019	269	60	22	3.8	5.9	0.5	0.5	3650
2020	259	55	25	3.7	6.1	0.4	0.6	3447
2021	329*	48	28	3.6	6.3	0.3	0.7	2836**
2022	309	62	30	3.9	6.4	0.2	0.8	3164
2023	299	70	32	4.0	6.5	0.1	0.9	3483
2024	279	75	34	4.1	6.5	0.0	1.0	3858
2025	269	80	35	4.2	6.6	0.1	1.1	3783

The complete dataset is:

Year & Price & Marketing & Competitors & Rating & Size & x_coord & y_coord & Sales

transformed data

For the year 2026, the planned features are specified as: Year: 2026; Price: transformed price USD; Marketing: transformed marketing thousand USD; Competitors: 37; Rating: 4.3; Size: 6.7 inches; x-coord: transformed x; y-coord: transformed y.

Perform these tasks: Forecast the sales volume for 2026 based on the dataset and the given 2026 features. Denote the predicted sales volume as hatY, and provide it in pieces. The design coordinates for the 2026 collection are hatx and haty. After processing, output only three numbers: the predicted hatY, followed by hatx, and then haty, each separated by a single space. Do not include any other text, explanations, or formatting in your response.

Prompt 2. *You are given a dataset of a handicraft company's pottery collection sales from 2016 to 2025. Each year in the dataset includes the following variables: Year index; Average retail price per piece (USD); Annual marketing expenditure (thousand USD); Number of direct competitor artisan brands; Average customer rating on e-commerce platforms (1-5 scale); Pottery size (inches, diameter); Design coordinate x (relative units); Design coordinate y (relative units); Annual sales volume (pieces).*

The complete dataset is:

Year & Price & Marketing & Competitors & Rating & Size & x_coord & y_coord & Sales

transformed data

For the year 2026, the planned features are specified as: Year: 2026; Price: transformed price USD; Marketing: transformed marketing thousand USD; Competitors: 37; Rating: 4.3; Size: 6.7 inches; x-coord: transformed x; y-coord: transformed y.

Perform these tasks: Estimate the 20th percentile of the sales volume for 2026, that is, a level that the actual sales volume is expected to exceed with probability 0.8, based on the dataset and the given 2026 features. Denote this value as Q20, and provide it in pieces. The design coordinates for the 2026 collection are hatx and haty. After processing, output only three numbers: Q20, followed by hatx, and then haty, each separated by a single space. Do not include any other text, explanations, or formatting in your response.

S4 Proofs of Lemmas and Theorems

Proof of Lemma 1. We verify the three subgroup axioms.

Identity: The identity map $\mathrm{id}_{\mathcal{X}}$ is globally traceable with recovery map $R_{\mathrm{id}} = \mathrm{id}_{\Theta}$, since trivially $A\{X\} \overset{d}{=} A\{X\}$ for every $X \in \mathcal{X}$.

Inverse: Suppose $t \in \mathcal{G}$ with associated recovery map R_t. Then for all $X \in \mathcal{X}$, $R_t\left(A\{t(X)\}\right) \overset{d}{=} A\{X\}$. Applying the homeomorphism R_t^{-1} to both arguments yields $A\{t(X)\} \overset{d}{=} R_t^{-1}\left(A\{X\}\right)$. Since t is a bijection, as X ranges over $\mathcal{X}$ the variable $Y = t(X)$ also ranges over $\mathcal{X}$. Rewriting the above with $X = t^{-1}(Y)$ yields $A\{Y\} \overset{d}{=} R_t^{-1}\left(A\{t^{-1}(Y)\}\right)$ for all $Y \in \mathcal{X}$. Hence t^{-1} is globally traceable with recovery map $R_{t^{-1}} = R_t^{-1}$, so $t^{-1} \in \mathcal{G}$.

Closure under composition: Let $t_1, t_2 \in \mathcal{G}$ with recovery maps R_{t_1}, R_{t_2}. For any $X \in \mathcal{X}$,

set $Y = t_1(X)$. Using the traceability of t_2 with input Y gives $R_{t_2}(A\{t_2(Y)\}) \overset{d}{=} A\{Y\}$. Applying the continuous map R_{t_1} to both arguments, we have $R_{t_1}(R_{t_2}(A\{t_2(t_1(X))\})) \overset{d}{=} R_{t_1}(A\{t_1(X)\})$. Because t_1 is also traceable, $R_{t_1}(A\{t_1(X)\}) \overset{d}{=} A\{X\}$. Combining the two equalities yields $(R_{t_1} \circ R_{t_2})(A\{(t_2 \circ t_1)(X)\}) \overset{d}{=} A\{X\}$. Thus $t_2 \circ t_1$ is globally traceable with recovery map $R_{t_2 \circ t_1} = R_{t_1} \circ R_{t_2}$, proving closure.

Since $\mathcal{G}$ contains the identity, is closed under inverses and composition, it is a subgroup of Homeo($\mathcal{X}$). □

Proof of Lemma 2. Since $X_1, X_2 \in GX_0$, there exist group elements $t_1, t_2 \in G$ such that $X_1 = t_1(X_0)$ and $X_2 = t_2(X_0)$. Define $g = t_2 \circ t_1^{-1} \in G$; then $X_2 = g(X_1)$.

To show local traceability, assume G is locally traceable at X_1. For any $t \in G$, we must show the condition holds at X_2. Using the local traceability at X_1, we have

$$R_g(R_t(A\{(t \circ g)(X_1)\})) \overset{d}{=} A\{X_1\} \overset{d}{=} R_g(A\{g(X_1)\})$$

and thus $R_t(A\{t(g(X_1))\}) \overset{d}{=} A\{g(X_1)\}$, meaning $R_t(A\{t(X_2)\}) \overset{d}{=} A\{X_2\}$. The converse direction follows by symmetry, swapping the roles of X_1 and X_2.

To show local transitivity and disclosure spaces, for any $X \in GX_0$, there exists some $h \in G$ with $X = h(X_0)$. The disclosure space at X is: $\Theta_{GX} = \{\, \mathbb{E}[A\{t^{-1}(X)\}] \mid t \in G \,\} = \{\, \mathbb{E}[A\{(t^{-1} \circ h)(X_0)\}] \mid t \in G \,\}$. As t runs over G, the composition $t^{-1} \circ h$ also runs over G (since G is a group). Therefore, $\Theta_{GX} = \{\, \mathbb{E}[A\{s^{-1}(X_0)\}] \mid s \in G \,\} = \Theta_{GX_0}$. This establishes that $\Theta_{GX_1} = \Theta_{GX_2} = \Theta_{GX_0}$. Consequently, if $\Theta_{GX_0} = \Theta$, then $\Theta_{GX} = \Theta$ for all $X \in GX_0$,

meaning G is locally transitive at every point in its orbit. □

Proof of Lemma 3. As a closed subgroup of Homeo($\mathcal{X}$), G inherits local compactness and the Hausdorff property from the ambient space. The group operations (composition and inversion) on G are restrictions of the continuous operations on Homeo($\mathcal{X}$) and thus remain continuous. Hence, G is a locally compact Hausdorff topological group. By Haar's theorem (Folland 2015), such a group admits a (left) Haar measure $\tilde{\mu}$, unique up to a positive scalar multiplier. □

Proof of Theorem 1. We prove each statement separately.

Size control under H_0. If G is locally traceable at X_0, then by Definition 3, $R_t\big(A\{t(X_0)\}\big) \overset{d}{=} A\{X_0\}$ for all $t \in G$. In particular, conditional on the i.i.d. samples $T_1, \ldots, T_n$ from μ, the random variables $R_{T_i}\big(A\{T_i(X_0)\}\big)$ all have the same distribution. Therefore, the processed data $\mathcal{Y}$ consist of n groups drawn from identical distributions. By Assumption 1, $\mathbb{P}\big(S(\mathcal{Y}) = 1 \mid \mathcal{T}\big) \leq \alpha$. Taking expectation over $\mathcal{T}$ yields $\mathbb{P}_{H_0}\big(S(\mathcal{Y}) = 1\big) \leq \alpha$.

Power under H_1. Under H_1, there exist open sets $U_1, U_2 \subset G$ such that any $t_1 \in U_1$ and $t_2 \in U_2$ yield different distributions. Because the Haar measure $\tilde{\mu}$ is strictly positive on non-empty open sets (Folland 2015), we have $\tilde{\mu}(U_1), \tilde{\mu}(U_2) > 0$, and thus $\mu(U_1), \mu(U_2) > 0$ as well. Consequently, as the number of sampled transformations n increases, the probability that $\mathcal{T}$ contains at least one element from each of U_1 and U_2 tends to one: $\mathbb{P}\big(\{\mathcal{T} \cap U_1 \neq \emptyset\} \cap \{\mathcal{T} \cap U_2 \neq \emptyset\}\big) \to 1$ as $n \to \infty$.

Now condition on the event that $\mathcal{T} \cap U_1 \neq \emptyset$ and $\mathcal{T} \cap U_2 \neq \emptyset$. Then among the n groups in

$\mathcal{Y}$, there exist at least two groups whose underlying distributions are different. By the power part of Assumption 1, the probability that S detects this difference tends to one as the per-group sample size m grows: $\mathbb{P}\big(S(\mathcal{Y}) = 1 \mid \{\mathcal{T} \cap U_1 \neq \emptyset\} \cap \{\mathcal{T} \cap U_2 \neq \emptyset\}\big) \to 1 \quad \text{as } m \to \infty$. Combining the two limits, we obtain $\mathbb{P}_{H_1}\big(S(\mathcal{Y}) = 1\big) \to 1$ as $m, n \to \infty$. This completes the proof. □

Bibliography

Arboretti, R., Barzizza, E., Biasetton, N. & Disegna, M. (2025), 'A review of multivariate permutation tests: Findings and trends', *Journal of Multivariate Analysis* **207**, 105421. **URL:** *https://www.sciencedirect.com/science/article/pii/S0047259X25000168*

Folland, G. B. (2015), *A Course in Abstract Harmonic Analysis*, 2nd edn, Chapman and Hall/CRC, New York. First published 2015, eBook published 3 February 2016.

Montgomery, D. C., Jennings, C. L. & Kulahci, M. (2024), *Introduction to Time Series Analysis and Forecasting*, WILEY SERIES IN PROBABILITY AND STATISTICS, third edn, Wiley.