WORKING PAPERS

N° 1769

September 2026

"Double Folding Test of Unimodality with an Application to Invariant Coordinate Selection"

Colombe Becquart, Aurore Archimbaud, Anne M. Ruiz and Zaineb Smida

Toulouse School of Economics

DOUBLE FOLDING TEST OF UNIMODALITY WITH AN APPLICATION TO INVARIANT COORDINATE SELECTION

A PREPRINT

Colombe Becquart *
Toulouse School of Economics,
Université Toulouse Capitole
France
Université de Toulouse
France

Aurore Archimbaud
TBS Business School
France

Anne M. Ruiz
Toulouse School of Economics,
Université Toulouse Capitole
France

Zaineb Smida
Univ Lyon, INSA Lyon, UJM, UCBL, ECL,
CNRS UMR 5208, ICJ
France

ABSTRACT

Testing for unimodality is of particular interest in clustering, where multimodality can be viewed as evidence of an underlying group structure. The Double Folding Test of Unimodality (DFTU) was recently proposed as an extension of the Folding Test of Unimodality (FTU), but its original bivariate test statistic does not yield a straightforward global p-value. In this paper, we introduce a new test statistic for the DFTU that admits a well-defined p-value. We derive its asymptotic distribution, leading to a fast approximation of the p-value that avoids computationally intensive Monte Carlo procedures. We then apply the resulting test to component selection in invariant coordinate selection (ICS), a dimension reduction method. The aim is to identify the subspace containing the relevant clustering information and thereby improve the performance of subsequent clustering procedures. In this setting, multimodality offers a natural criterion for identifying informative components, as an alternative to the commonly used non-normality criterion. Applications show that the proposed unimodality-based criterion outperforms the normality-based criterion in identifying the subspace of interest in clustering settings.

1 Introduction

Dimension reduction is often used to improve cluster identification, an approach commonly referred to as tandem clustering (Arabie and Hubert, 1994). Common dimension reduction techniques in this context include principal component analysis (PCA) (Hotelling, 1933; Jolliffe, 2002), and invariant coordinate selection (ICS) (Tyler et al., 2009; Alfons et al., 2024). A well-known limitation of PCA-based approaches is that component selection is typically driven by variance, which does not necessarily preserve cluster structure. Although ICS is better suited to recover informative directions, the selection of relevant components remains challenging. It is often determined by non-normality, assessed through normality tests applied to the components (Alfons et al., 2024). This approach, referred to as the normal criterion, retains only the components showing significant evidence of non-normality because it assumes that departures from normality indicate the presence of relevant information, whereas normal components are considered noise. This criterion is particularly common in projection pursuit (Huber, 1985; Friedman, 1987), but other criteria may be considered.

*Corresponding author: colombe.becquart@tse-fr.eu
Code and replication files are available at `https://github.com/cbecquart/DFTU_ICS-Replication`.

As discussed in Becquart et al. (2026), unimodality tests offer a promising approach to selecting invariant components in a clustering context. In this setting, the modes can be interpreted as proxies for the clusters. Consequently, unimodality indicates the absence of relevant clustering information, and the focus should be placed on the subspace exhibiting multimodality. This provides an alternative criterion for ICS component selection, based on multimodality rather than non-normality.

Following the methodology of the normal criterion (Alfons et al., 2024), a simple procedure consists of applying a univariate test to each invariant coordinate in a predefined order. The tests are applied sequentially until the null hypothesis is no longer rejected, corresponding to normality in Alfons et al. (2024) and unimodality in our case. Decisions are based on p-values that are adjusted for multiple testing.

The main approaches to testing unimodality, developed in the 1980s, are Hartigan's dip test (Hartigan and Hartigan, 1985), based on the cumulative distribution function, and Silverman's test (Silverman, 1981), based on kernel density estimation. More recently, Siffer et al. (2018) introduced an alternative approach aimed at computational efficiency and multivariate extension. The resulting Folding Test of Unimodality (FTU) is based on a variance reduction criterion. Becquart et al. (2026) analyzed a limitation of the FTU and proposed a two-step enhancement in the univariate case, leading to the Double Folding Test of Unimodality (DFTU). In the present paper, we go one step further by proposing a new test statistic for the DFTU and a corresponding procedure for computing its p-value.

While the DFTU showed promising results, its bivariate statistic (Φ_{1n}, Φ_{2n}), whose formal definition is recalled in Section 2, does not admit a straightforward computation of the associated p-value. The overall significance level α is decomposed into two parts: α_1, associated with Φ_{1n}, and α_2, associated with Φ_{2n}, such that

$$\alpha = \alpha_1 + (1 - \alpha_1)\alpha_2.$$

This decomposition corresponds to

$$\alpha = P(\Phi_{1n} < q_1) + P(\Phi_{1n} \geq q_1)\, P(\Phi_{2n} < q_2 \mid \Phi_{1n} \geq q_1),$$

where q_1 and q_2 are the critical values corresponding to significance levels α_1 and α_2, respectively.

To compute the associated p-value, two issues arise. First, the p-value should not depend on the significance level α (and thus neither on α_1 nor α_2), whereas the algorithm requires a threshold q_1 that depends on α_1. Second, denoting by x_1 and x_2 the observed values of Φ_{1n} and Φ_{2n}, it is not meaningful to compare $\alpha_1 + (1 - \alpha_1)\alpha_2$ with

$$P(\Phi_{1n} < x_1) + P(\Phi_{1n} \geq x_1)\, P(\Phi_{2n} < x_2 \mid \Phi_{1n} \geq x_1),$$

(or, equivalently, with a version conditioned on q_1). Indeed, if the algorithm stops at the first step, the value of x_2 from the second step is not taken into account. For example, folding a symmetric two-Gaussian mixture results in a unimodal distribution, which can yield a large value of x_2. Conversely, if the procedure proceeds to the second step, this indicates that the first step may have failed. In that case, the value of x_1 should not be considered reliable when x_2 is small.

To overcome this difficulty, Section 2 introduces a new test statistic for the DFTU. Unlike the original formulation, this statistic admits a well-defined p-value. Moreover, the derivation of its asymptotic distribution provides a fast approximation of this p-value, avoiding the need for computationally intensive Monte Carlo procedures. Section 3 then presents the application to ICS component selection, and Section 4 concludes.

2 Revisiting the DFTU with a new univariate test statistic

2.1 Definition of the new test statistic

We begin by recalling the construction of the DFTU. Let X be a real-valued random variable with finite fourth moment. Let

$$s_1^* = \underset{s \in \mathbb{R}}{\operatorname{argmin}}\, \mathrm{Var}|X - s|, \qquad s_1^{**} = \underset{s \in \mathbb{R}}{\operatorname{argmin}}\, \mathrm{Var}[(X - s)^2],$$

be the exact and approximate pivots introduced in Siffer et al. (2018).

Using these pivots, construct the folded variables

$$Y^* = |X - s_1^*|, \qquad Y^{**} = |X - s_1^{**}|.$$

The second folding step is obtained by applying the same procedure to the variable Y^{**}. Namely, let

$$s_2^* = \underset{s \in \mathbb{R}}{\operatorname{argmin}}\, \mathrm{Var}|Y^{**} - s|,$$

and set

$$Z = |Y^{**} - s_2^*|.$$

Finally, define the parameters

$$\Phi_1 = \frac{4\text{Var}[Y^*]}{\text{Var}[X]}, \qquad \Phi_2 = \frac{4\text{Var}[Z]}{\text{Var}[Y^{**}]}.$$

The hypotheses considered in Becquart et al. (2026) are

$$\mathcal{H}_0 : \Phi_1 \geq 1 \text{ and } \Phi_2 \geq 1 \quad \text{vs.} \quad \mathcal{H}_1 : \Phi_1 < 1 \text{ or } \Phi_2 < 1.$$

The null hypothesis is based on a new definition of unimodality, which considers a distribution to be unimodal when $\Phi_1 \geq 1$ and $\Phi_2 \geq 1$. This definition builds upon the work of Siffer et al. (2018). These hypotheses can be equivalently written as

$$\mathcal{H}_0 : \min(\Phi_1, \Phi_2) \geq 1 \quad \text{vs.} \quad \mathcal{H}_1 : \min(\Phi_1, \Phi_2) < 1.$$

Let $X_1, \dots, X_n$ be an i.i.d. sample from the distribution of X. Let s_{1n}^*, s_{1n}^{**} and s_{2n}^* denote the empirical counterparts of s_1^*, s_1^{**}, and s_2^* computed using the unbiased variance estimator. On the sample $X_1, \dots, X_n$, it is given by:

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2, \qquad \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

The corresponding transformed variables are

$$Y_i^* = |X_i - s_{1n}^*|, \qquad Y_i^{**} = |X_i - s_{1n}^{**}|, \qquad Z_i = |Y_i^{**} - s_{2n}^*|.$$

Denote by $S_{Y^*}^2$, $S_{Y^{**}}^2$, and S_Z^2 the unbiased variance estimators computed from $(Y_i^*)_{i=1}^n$, $(Y_i^{**})_{i=1}^n$, and $(Z_i)_{i=1}^n$, respectively. Finally, define

$$\Phi_{1n} = \frac{4S_{Y^*}^2}{S_X^2}, \qquad \Phi_{2n} = \frac{4S_Z^2}{S_{Y^{**}}^2},$$

and the test statistic

$$T_n = \min(\Phi_{1n}, \Phi_{2n}).$$

To compute the p-value, a unimodal distribution must be specified as the null distribution. Following Hartigan and Hartigan (1985), Siffer et al. (2018) and Becquart et al. (2026), we use the uniform distribution, as it represents a boundary case between unimodality and multimodality.

Remark 1. *The choice of the uniform distribution as the reference unimodal distribution is also used in the construction of the parameters Φ_1 and Φ_2. They are defined as ratios between the variance reduction obtained by folding the distribution and the corresponding variance reduction for the reference uniform distribution, which is equal to $1/4$ in both cases. See Siffer et al. (2018) and Becquart et al. (2026) for details.*

Under the uniform null distribution, denoted by $\mathcal{U}$, the p-value associated with an observed value $T_n = t$ is

$$p_n = P_{\mathcal{U}}(T_n \leq t).$$

For simplicity, we write P instead of $P_{\mathcal{U}}$. We then have

$$\begin{aligned}
p_n &= 1 - P(T_n > t) \\
&= 1 - P(\Phi_{1n} > t, \Phi_{2n} > t) \\
&= P(\Phi_{1n} \leq t) + P(\Phi_{1n} > t)\, P(\Phi_{2n} \leq t \mid \Phi_{1n} > t) \\
&= P(\Phi_{1n} \leq t) + P(\Phi_{2n} \leq t) - P(\Phi_{1n} \leq t)\, P(\Phi_{2n} \leq t \mid \Phi_{1n} \leq t) \\
&= P(\Phi_{1n} \leq t) + P(\Phi_{2n} \leq t) - P(\Phi_{1n} \leq t, \Phi_{2n} \leq t).
\end{aligned}$$

The p-value is derived from the joint null distribution of (Φ_{1n}, Φ_{2n}). Since this distribution is unknown, we derive its asymptotic approximation in Proposition 1. The main steps of the proof, presented in Appendix A, assume that the effect of estimating the empirical pivots on the asymptotic distribution is negligible.

Proposition 1. *Let*

$$\mathbf{\Phi}_n = (\Phi_{1n}, \Phi_{2n})^\top.$$

Under H_0,

$$\sqrt{n} \left(\mathbf{\Phi}_n - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \xrightarrow{d} \mathcal{N}(0, \Sigma_\Phi),$$

where

$$\Sigma_\Phi = \begin{pmatrix} \frac{6}{5} & -\frac{9}{20} \\ -\frac{9}{20} & \frac{6}{5} \end{pmatrix}.$$

2.2 Implementation

From Section 2.1, the p-value is given by

$$p_n = P(\Phi_{1n} \leq t) + P(\Phi_{2n} \leq t) - P(\Phi_{1n} \leq t, \Phi_{2n} \leq t).$$

Using the asymptotic results, we can approximate the p-value by

$$p_n \approx 2\, F_{\mathcal{N}(1,\sigma^2)}(t) - F_{\mathcal{N}_2(\mu,\Sigma_\Phi)}(t,t).$$

where $\sigma^2 = \frac{6}{5}$ and $\Sigma_\Phi = \begin{pmatrix} \frac{6}{5} & -\frac{9}{20} \\ -\frac{9}{20} & \frac{6}{5} \end{pmatrix}$.

Simulations show that this approximation closely matches the Monte Carlo estimate for various values of t and n, as illustrated in Figure 1.

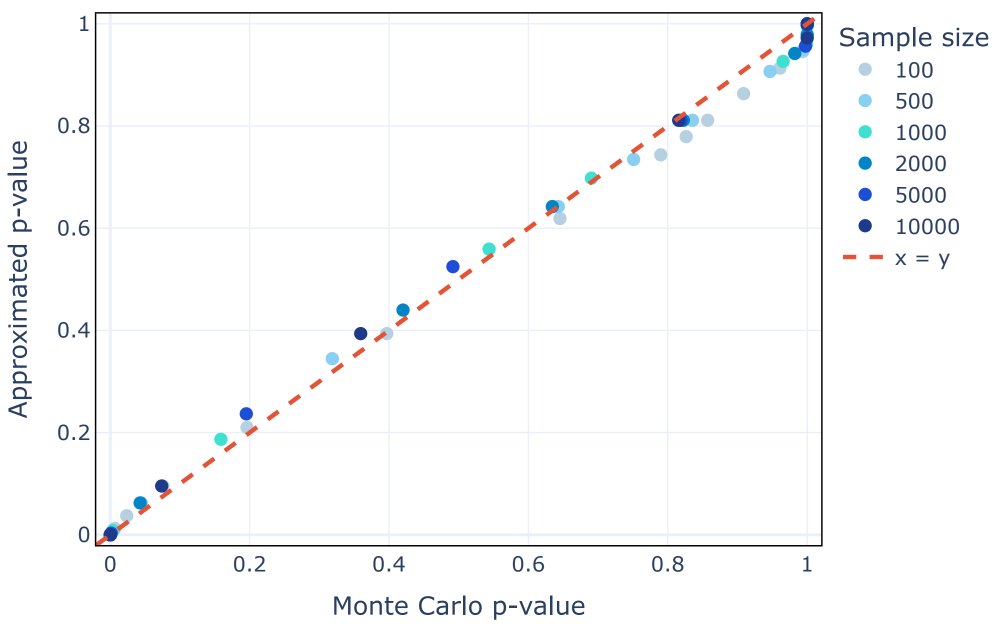


Figure 1: Comparison between the approximation formula and Monte Carlo estimation of the p-value.

Appendix B provides the p-values obtained with this approximation for the simulations from Becquart et al. (2026), along with two additional examples.

3 Application to ICS component selection

3.1 Testing procedure

As stated in the introduction, the unimodality-based criterion considers that relevant invariant coordinates (Z_j), i.e., the projected data along the invariant directions, exhibit multimodal structure whereas noise coordinates are unimodal. The selection method described in this section relies on the DFTU to assess if coordinates are significantly multimodal or not. It follows the same testing procedure as the normal criterion presented in Alfons et al. (2024)[2] (see also Archimbaud et al., 2018).

[2]Following the machine learning literature, we use the term "components" to refer to the eigenvectors. In Alfons et al. (2024), the term is instead used as a synonym for coordinates.

Algorithm 1 Unimodality-based component selection

Require: invariant coordinates Z, level $\alpha \in (0, 1)$, max components $m_{\max}$
Ensure: selected components $\mathcal{S}$
$\mathcal{C} \leftarrow \{IC_1, \ldots, IC_p\}$ ▷ Set of invariant components
$\mathcal{S} \leftarrow \emptyset$ ▷ Set of selected components
$m \leftarrow 0$ ▷ Number of selected components
$pv_j \leftarrow \mathrm{DFTU}(Z_j)$ for $j = 1, \ldots, p$ ▷ Compute p-values
$pv \leftarrow \{pv_1, \ldots, pv_p\}$
if $\min_j pv_j > \alpha$ **then**
  return $\mathcal{S}$
else
  while $m \leq m_{\max}$ **do**
    $pv_L \leftarrow pv_1$, $pv_R \leftarrow pv_{p-m}$
    if $pv_L < \alpha$ **and** $pv_L < pv_R$ **then**
      $\mathcal{S} \leftarrow \mathcal{S} \cup \{C_1\}$
      Remove first element from $\mathcal{C}$ and from pv
    else if $pv_R < \alpha$ **then**
      $\mathcal{S} \leftarrow \mathcal{S} \cup \{C_{p-m}\}$
      Remove last element from $\mathcal{C}$ and from pv
    else
      break
    end if
    $m \leftarrow m + 1$
    $\alpha \leftarrow \alpha\, m/(m + 1)$
  end while
  return $\mathcal{S}$
end if

The relevant invariant directions are either the first ones (associated with large generated kurtosis) or the last ones (associated with small generated kurtosis) (Tyler et al., 2009, Theorem 4). As outlined in Algorithm 1, the procedure takes the invariant coordinates, the significance level α, and a maximum number of directions to select as input. It then applies the DFTU to each coordinate, using the approximation formula from Section 2.2 to obtain marginal p-values. Although multimodality is assessed on coordinates, the selected objects are the associated invariant components, i.e., the invariant directions defining these coordinates. Components are selected iteratively from an initial set containing all components by examining the p-values computed on the first and last coordinates. At each iteration, the component whose coordinate provides the strongest evidence against unimodality (i.e., the smallest p-value below the significance level) is selected and removed from the set. The significance level is then updated to account for multiple testing by dividing the original significance level α by the current iteration count. This procedure continues until no further components meet the selection criteria or the predefined maximum number of components is reached.

3.2 Real data application

In this section, we apply the component selection procedure to real datasets and compare it with normality tests. As the unimodality-based procedure is only relevant in a clustering context, our application builds upon the examples of Alfons et al. (2024) who examine ICS for clustering. For each of the three datasets, they provide the best-performing scatter pair for computing ICS, the corresponding target set of components to retain, and the results of the normality tests. Specifically, they use the D'Agostino test for normality based on skewness (D'Agostino et al., 1990). In this analysis, we additionally apply the unimodality-based selection procedure to the invariant components obtained from the best scatter pair.

3.2.1 Iris

The *iris* dataset (Anderson, 1936; Fisher, 1936) contains 150 observations from three iris species, each representing one third of the dataset. The samples are characterized by four numerical features. Figure 2a displays the resulting invariant coordinates, where colors are used to distinguish the underlying species, although this label information was not used during the unsupervised ICS computation. The first invariant component clearly appears to be the relevant one, as it provides a separation between the three species. Figure 2b shows the corresponding coordinates without the cluster color coding. According to Alfons et al. (2024), the normality-based criterion selects only the first component, as it is the only one identified as non-Gaussian by the D'Agostino skewness test. The unimodal criterion reaches the same conclusion: only the first coordinate exhibits a multimodal structure, indicating that the first invariant component should be retained. Both selection methods are successful in this example.

3.2.2 Crabs

The *crabs* dataset (Campbell and Mahon, 1974) gathers morphological measurements of crabs. With 200 observations on 5 variables, it comprises 4 clusters corresponding to the two sexes of two species. Each cluster represents 25% of the observations. Following Alfons et al. (2024), the data are log-transformed before the analysis. The target components are IC1 and IC2. As shown in Figure 3a, these two components are sufficient to identify all four clusters. The normal criterion does not select any components, whereas the unimodality-based procedure correctly identifies the targeted subspace by selecting the first two components. As illustrated in Figure 3b, the first two invariant coordinates clearly exhibit a bimodal structure. Nevertheless, they are still regarded as Gaussian by the D'Agostino test, explaining why this test fails to identify them as relevant clustering directions.

3.2.3 Philips

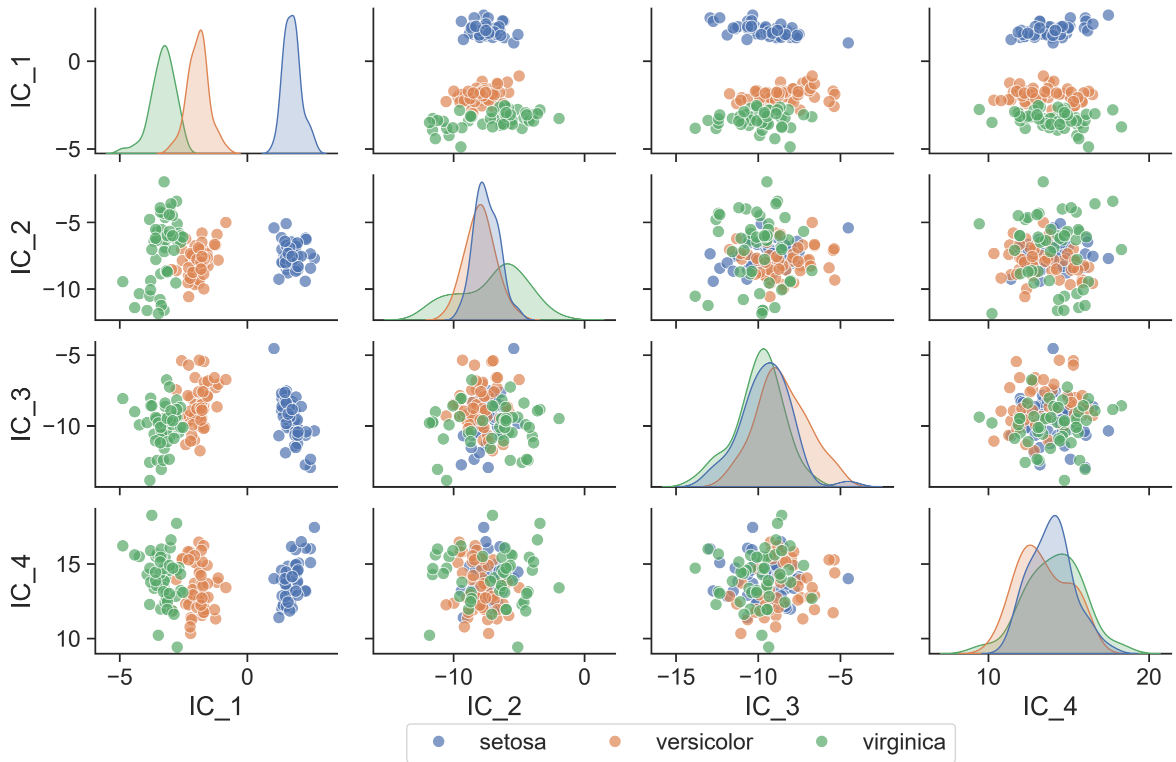
The *Philips* dataset (Rousseeuw and Driessen, 1999) contains 677 observations of television sets described by nine features. It was originally introduced for anomaly detection studies, and Philips engineers identified three groups: one main cluster containing 74% of the observations and two outlier groups representing 15% and 11% of the data, respectively. As shown in Figure 4a, the target components are the first two. The normal criterion selects the first four components, thus overestimating the relevant subspace, whereas the unimodal criterion retains only the first component, missing the second one. As illustrated in Figure 4b, the first coordinate is clearly bimodal. In contrast, the second coordinate is classified as unimodal because Group 2 (shown in blue) is relatively small (11% of the observations) and is not well separated from the other groups.

4 Conclusion

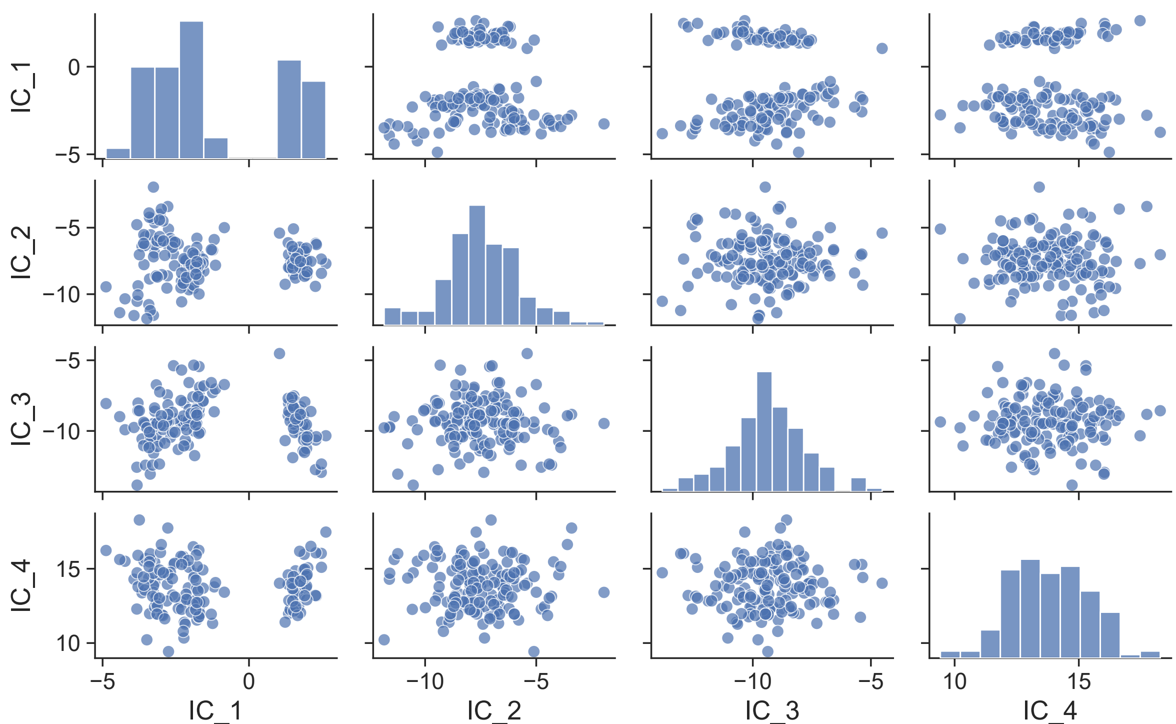
In this work, we introduced a novel test statistic for the DFTU that enables the direct computation of p-values. To bypass the computational cost of Monte Carlo simulations, we derived an approximation formula grounded in the statistic's asymptotic behavior. Additionally, we proposed a new testing procedure for selecting ICS components based on a unimodal criterion and the DFTU. It shows promising results, based on the real data application. Both the enhanced DFTU and the ICS component selection procedure have been implemented in Python, and made publicly available in the `ICSpyLab` package (Becquart and Abdelsameia, 2026; Becquart, 2026). As for future directions, extending the DFTU to a bivariate setting could improve ICS applications by enabling the detection of modes that are not marginally detectable. It would also be interesting to conduct a broader comparative evaluation with the normal criterion.

Acknowledgments

Anne M. Ruiz acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010.

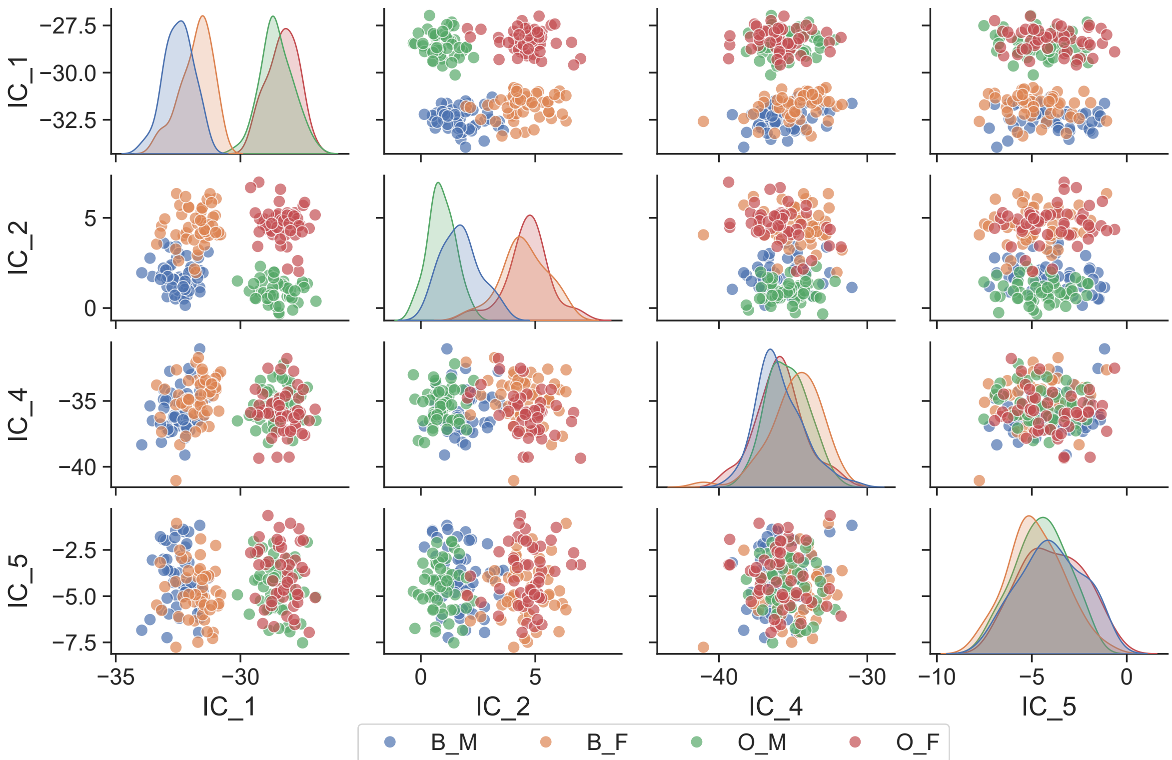


(a) Groups shown in color

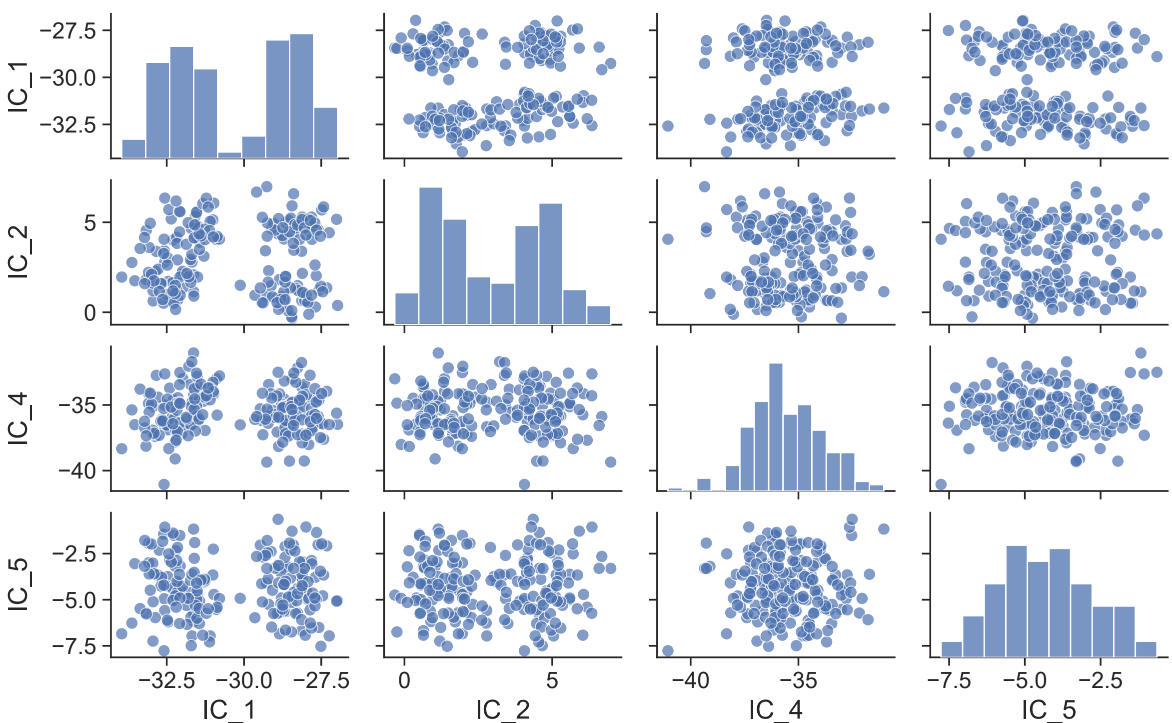


(b) Groups not shown

Figure 2: Scatterplots of the first and last invariant coordinates for the *iris* dataset, with groups shown in color (a) and without group labels (b).

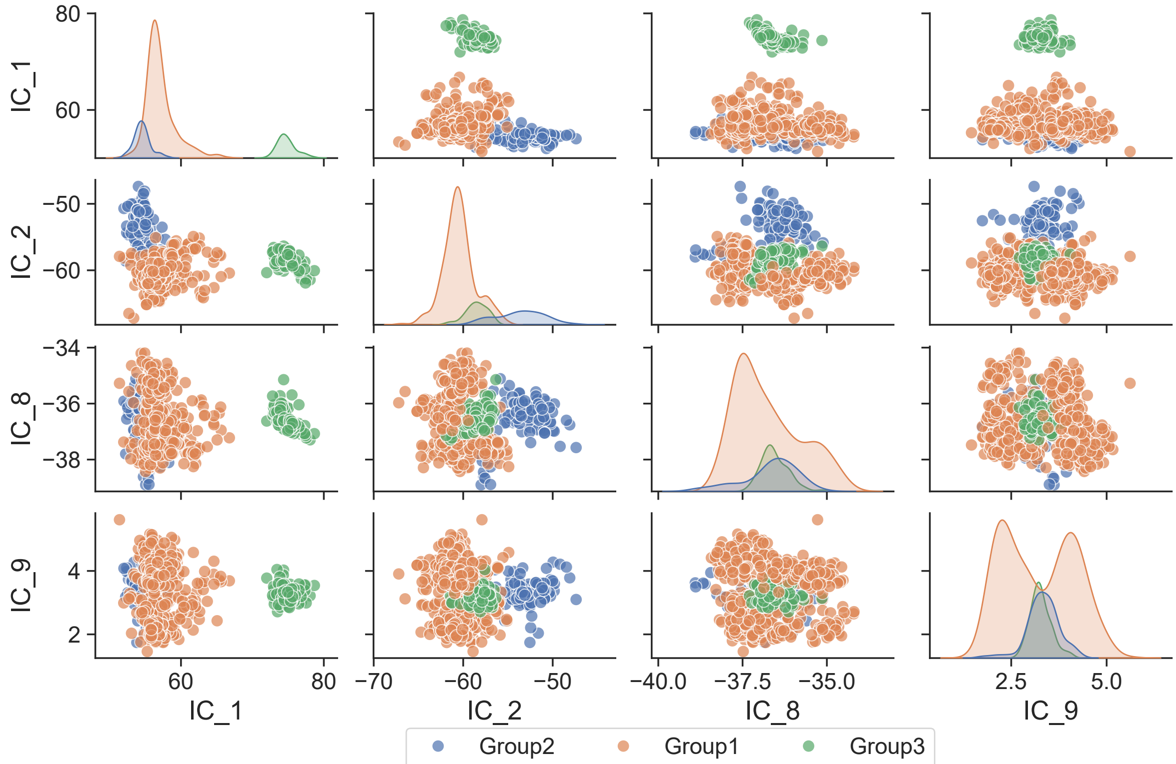


(a) Groups shown in color

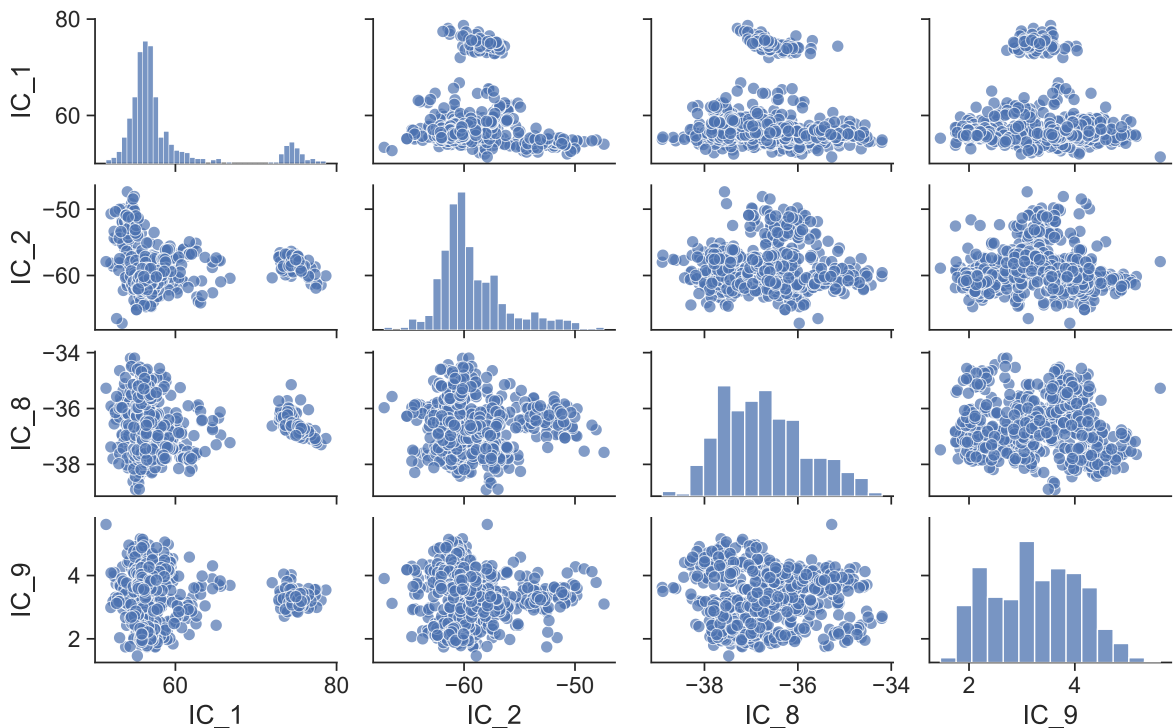


(b) Groups not shown

Figure 3: Scatterplots of the first and last invariant coordinates for the *crabs* dataset, with groups shown in color (a) and without group labels (b).



(a) Groups shown in color



(b) Groups not shown

Figure 4: Scatterplots of the first and last invariant coordinates for the *Philips* dataset, with groups shown in color (a) and without group labels (b).

A Proof of Proposition 1

Lemma A.0.1. *Let* $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{U}([0,1])$*, and define*

$$Y_i = |X_i - \mathbb{E}[X_1]|\,, \qquad Z_i = |Y_i - \mathbb{E}[Y_1]|\,, \qquad i = 1, \dots, n.$$

Then

$$Y_1, \dots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{U}([0, \mathbb{E}[X_1]])\,, \qquad Z_1, \dots, Z_n \overset{\text{i.i.d.}}{\sim} \mathcal{U}([0, \mathbb{E}[Y_1]])\,.$$

Moreover, for

$$W_i = \left(X_i, X_i^2, Y_i, Y_i^2, Z_i, Z_i^2\right)^\top \in \mathbb{R}^6,$$

the vectors $W_1, \dots, W_n$ *are i.i.d. with finite second moments. Hence, by the multivariate Central Limit Theorem,*

$$\sqrt{n}\left(\bar{W}_n - \mathbb{E}[W_1]\right) \xrightarrow{d} \mathcal{N}(0, \Omega),$$

where

$$\bar{W}_n = \frac{1}{n}\sum_{i=1}^{n} W_i, \qquad \Omega = \mathrm{Cov}(W_1).$$

Proof. Let $X \sim \mathcal{U}([0,1])$ and define

$$Y = |X - \mathbb{E}[X]|\,, \quad \text{where} \quad \mathbb{E}[X] = \frac{1}{2}.$$

For $0 \le y \le \mathbb{E}[X]$,

$$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(\mathbb{E}[X] - y \le X \le \mathbb{E}[X] + y) = 2y,$$

which is the cumulative distribution function of the uniform distribution on $[0, \mathbb{E}[X]]$. Hence,

$$Y \sim \mathcal{U}([0, \mathbb{E}[X]])\,.$$

Applying the same argument to

$$Z = |Y - \mathbb{E}[Y]| = ||X - \mathbb{E}[X]| - \mathbb{E}[Y]|\,,$$

with $Y \sim \mathcal{U}([0, \mathbb{E}[X]])$, $\mathbb{E}[Y] = \frac{1}{4}$, yields

$$Z \sim \mathcal{U}([0, \mathbb{E}[Y]])\,.$$

Since Y_i and Z_i are measurable transformations of the i.i.d. variables X_i, the sequences (Y_i) and (Z_i) are themselves i.i.d. with the stated distributions.

Finally, each component of W_i is bounded, so W_i has finite second moments. Therefore, the multivariate Central Limit Theorem applies, giving

$$\sqrt{n}\left(\bar{W}_n - \mathbb{E}[W_1]\right) \xrightarrow{d} \mathcal{N}(0, \Omega),$$

where $\Omega = \mathrm{Cov}(W_1)$.

□

Lemma A.0.2. *Let* $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{U}([0,1])$*, and define*

$$Y_i = \left|X_i - \frac{1}{2}\right|, \qquad Z_i = \left|Y_i - \frac{1}{4}\right|, \qquad i = 1, \dots, n.$$

Let

$$W_i = \left(X_i, X_i^2, Y_i, Y_i^2, Z_i, Z_i^2\right)^\top \in \mathbb{R}^6, \qquad \bar{W}_n = \frac{1}{n}\sum_{i=1}^{n} W_i.$$

Denote by S_X^2, S_Y^2, S_Z^2 *the sample variances of* $(X_1, \dots, X_n)$, $(Y_1, \dots, Y_n)$*, and* $(Z_1, \dots, Z_n)$*, respectively. Define*

$$\Phi_{1n} = \frac{4S_Y^2}{S_X^2}, \qquad \Phi_{2n} = \frac{4S_Z^2}{S_Y^2}.$$

Then,

$$\sqrt{n}\begin{pmatrix} \Phi_{1n} - 1 \\ \Phi_{2n} - 1 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma_\Phi),$$

where

$$\Sigma_\Phi = \begin{pmatrix} \frac{6}{5} & -\frac{9}{20} \\ -\frac{9}{20} & \frac{6}{5} \end{pmatrix}.$$

Proof. By Lemma A.0.1, we have

$$\sqrt{n}\left(\bar{W}_n - \mu\right) \xrightarrow{d} \mathcal{N}(0, \Omega),$$

where $\mu = \mathbb{E}[W_1]$, $\Omega = \mathrm{Cov}(W_1)$.

We first apply the multivariate Delta method to the function $g : \mathbb{R}^6 \to \mathbb{R}^3$ defined by

$$g(m_1, m_2, m_3, m_4, m_5, m_6) = \begin{pmatrix} m_2 - m_1^2 \\ m_4 - m_3^2 \\ m_6 - m_5^2 \end{pmatrix}.$$

For simplicity of notation, we divide the sample variances by n rather than $n-1$. Evaluating g at $\bar{W}_n$ and at $\mu = \mathbb{E}[W_1]$ gives respectively

$$g(\bar{W}_n) = (S_X^2, S_Y^2, S_Z^2)^\top, \qquad g(\mu) = (\sigma_X^2, \sigma_Y^2, \sigma_Z^2)^\top,$$

where $\sigma_X^2 = \mathrm{Var}(X_1)$, $\sigma_Y^2 = \mathrm{Var}(Y_1)$, and $\sigma_Z^2 = \mathrm{Var}(Z_1)$.

Since g is continuously differentiable, the Delta method yields

$$\sqrt{n}\left(\begin{pmatrix} S_X^2 \\ S_Y^2 \\ S_Z^2 \end{pmatrix} - \begin{pmatrix} \sigma_X^2 \\ \sigma_Y^2 \\ \sigma_Z^2 \end{pmatrix}\right) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where $\Sigma = Dg(\mu)\Omega Dg(\mu)^\top$. The proof is stated using the sample variances with denominator n. The unbiased version is obtained by multiplying the sample variances by $n/(n-1)$. Since $n/(n-1) \longrightarrow 1$, the same asymptotic result holds for the unbiased sample variances used in the paper.

The Jacobian matrix of g evaluated at $m = (m_1, \dots, m_6)^\top \in \mathbb{R}^6$ is

$$Dg(m) = \begin{pmatrix} -2m_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2m_3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2m_5 & 1 \end{pmatrix}.$$

For the variables considered here,

$$\sigma_X^2 = \frac{1}{12}, \qquad \sigma_Y^2 = \frac{1}{48}, \qquad \sigma_Z^2 = \frac{1}{192}.$$

The diagonal terms of Σ are given by the definition of the asymptotic variance of the empirical variance:

$$\Sigma_{11} = \mu_{4X} - \sigma_X^4$$

where $\mu_{4X} = \mathbb{E}\left[(X - \mathbb{E}[X])^4\right]$ and $\sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$.

Since $X \sim U([0,1])$, we have $\sigma_X^2 = \frac{1}{12}$ and $\mu_{4X} = \frac{1}{80}$, which yields $\Sigma_{11} = \frac{1}{180}$. Similarly, the remaining diagonal elements are $\Sigma_{22} = \frac{1}{2880}$ and $\Sigma_{33} = \frac{1}{46080}$.

Regarding the off-diagonal elements, the first covariance term is given by:

$$\begin{aligned} \Sigma_{12} = \mathrm{Cov}(S_X^2, S_Y^2) &= \nabla g_1^T \Omega \nabla g_2 = \mathrm{Cov}\left(\nabla g_1^T W, \nabla g_2^T W\right) \\ &= \mathrm{Cov}\left((X - \mu_X)^2, (Y - \mu_Y)^2\right) \\ &= \mathbb{E}[(X - \mu_X)^2 (Y - \mu_Y)^2] - \sigma_X^2 \sigma_Y^2 \\ &= \frac{1}{2880} \end{aligned}$$

Similarly, $\Sigma_{23} = \Sigma_{13} = \frac{1}{46080}$. The covariance matrix of the sample variances is then

$$\Sigma = \begin{pmatrix} \frac{1}{180} & \frac{1}{2880} & \frac{1}{46080} \\ \frac{1}{2880} & \frac{1}{2880} & \frac{1}{46080} \\ \frac{1}{46080} & \frac{1}{46080} & \frac{1}{46080} \end{pmatrix}.$$

Next, define the function $h : \mathbb{R}^3 \to \mathbb{R}^2$ by

$$h(v_1, v_2, v_3) = \begin{pmatrix} \dfrac{4v_2}{v_1} \\ \dfrac{4v_3}{v_2} \end{pmatrix}.$$

By construction, $h(S_X^2, S_Y^2, S_Z^2) = (\Phi_{1n}, \Phi_{2n})^\top$. Applying the multivariate Delta method once again yields

$$\sqrt{n}\left(h(S_X^2, S_Y^2, S_Z^2) - h(\sigma_X^2, \sigma_Y^2, \sigma_Z^2)\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_\Phi),$$

where

$$\Sigma_\Phi = Dh(\sigma_X^2, \sigma_Y^2, \sigma_Z^2)\, \Sigma\, Dh(\sigma_X^2, \sigma_Y^2, \sigma_Z^2)^\top,$$

with $Dh(v)$ the Jacobian matrix of h evaluated at $v = (v_1, v_2, v_3)^\top \in \mathbb{R}^3$. A direct computation gives

$$\Sigma_\Phi = \begin{pmatrix} \frac{6}{5} & -\frac{9}{20} \\ -\frac{9}{20} & \frac{6}{5} \end{pmatrix}, \qquad h(\sigma_X^2, \sigma_Y^2, \sigma_Z^2) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Therefore,

$$\sqrt{n}\begin{pmatrix} \Phi_{1n} - 1 \\ \Phi_{2n} - 1 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma_\Phi),$$

which proves the result. □

Since the test statistic is affine equivariant (Becquart et al., 2026, see Proposition A.4.1), we may assume without loss of generality that, under the null hypothesis,

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{U}([0,1]).$$

Based on the notation introduced in Section 2.1, define the random vector

$$W_i = \left(X_i, X_i^2, Y_i^*, Y_i^{*2}, Y_i^{**}, Y_i^{**2}, Z_i, Z_i^2\right)^\top \in \mathbb{R}^8, \qquad \bar{W}_n = \frac{1}{n}\sum_{i=1}^n W_i.$$

Recall the statistics

$$\Phi_{1n} = \frac{4S_{Y^*}^2}{S_X^2}, \qquad \Phi_{2n} = \frac{4S_Z^2}{S_{Y^{**}}^2},$$

where S_X^2, $S_{Y^*}^2$, $S_{Y^{**}}^2$, and S_Z^2 denote the sample variances computed from $(X_i)_{i=1}^n$, $(Y_i^*)_{i=1}^n$, $(Y_i^{**})_{i=1}^n$, and $(Z_i)_{i=1}^n$, respectively.

The population counterparts of the estimators s_{1n}^*, s_{1n}^{**}, and s_{2n}^* are $\frac{1}{2}$, $\frac{1}{2}$, and $\frac{1}{4}$, respectively. This result is derived from Proposition A.3.1 of Becquart et al. (2026). The asymptotic distribution of (Φ_{1n}, Φ_{2n}) follows directly from Lemma A.0.2. In practice, the computation of (Φ_{1n}, Φ_{2n}) requires estimating s_{1n}^*, s_{1n}^{**}, and s_{2n}^*. However, the simulation results in Figure 1 suggest that the impact of estimating these quantities is negligible.

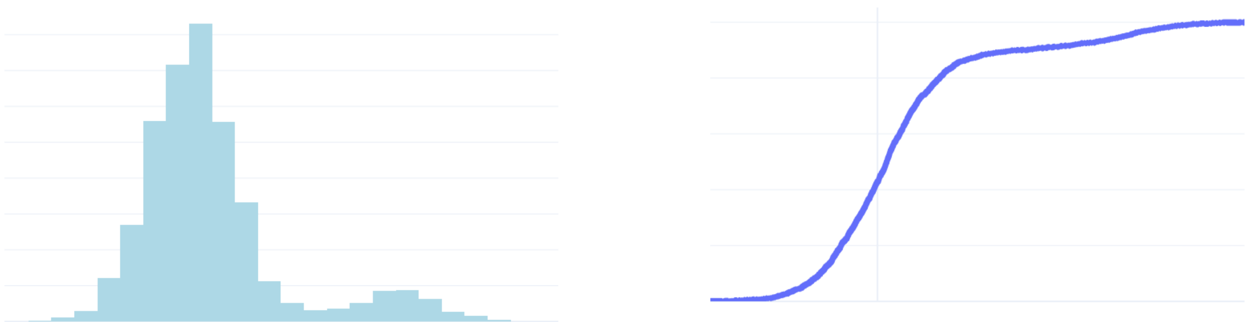
B Simulations

In this section, we reproduce the simulations from Becquart et al. (2026) using the novel implementation of the DFTU which uses the asymptotic approximation formula for p-value computation.

For each distribution, 100 datasets of $n = 1000$ observations were generated. Table 1 reports the average p-values across repetitions and the associated decision (unimodality versus multimodality) for the DFTU and Hartigan's dip test (Hartigan and Hartigan, 1985). The decisions are identical to those reported in Becquart et al. (2026). Table 1 also displays results for two additional distributions, illustrated in Figures 5 and 6. These additional examples demonstrate that when a group proportion is too small and no gap separates the groups, unimodality is not an adequate criterion for detecting group structure, as both tests conclude unimodality.

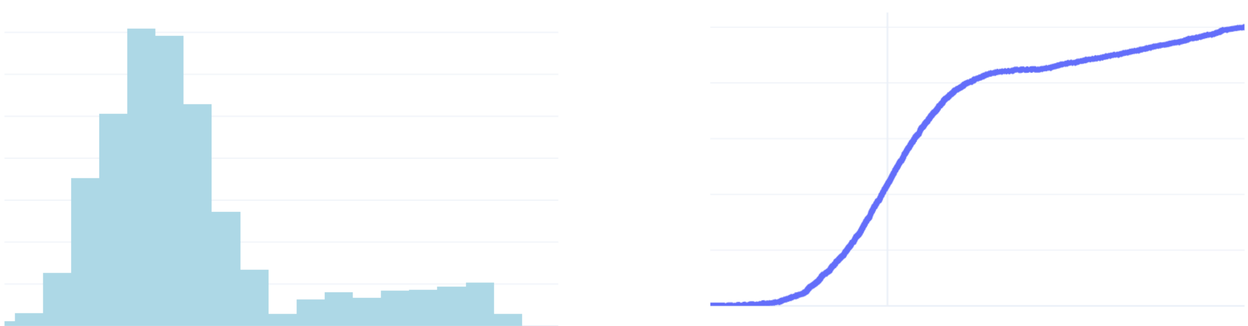
Distribution	DFTU		dip test	
$\mathcal{N}(0,1)$	1	Uni	0.93	Uni
$0.5 \cdot (\mathcal{N}_0 + \mathcal{N}_1)$	1	Uni	0.85	Uni
$0.6 \cdot \mathcal{N}_0 + 0.4 \cdot \mathcal{U}([4,8])$	0	Multi	0	Multi
$0.6 \cdot \mathcal{N}_0 + 0.4 \cdot \mathcal{U}([1,4])$	0	Multi	0.75	Uni
$1/3 \cdot \delta_{-2} + 1/3 \cdot \delta_0 + 1/3 \cdot \delta_2$	0	Multi	0	Multi
$1/3 \cdot \mathcal{N}_{-2} + 1/3 \cdot \mathcal{N}_0 + 1/3 \cdot \mathcal{N}_2$	0	Multi	0	Multi
$0.2 \cdot (\delta_{-3} + \delta_{-1.5} + \delta_{2.5} + \delta_4 + \delta_{11})$	0	Multi	0	Multi
$0.2 \cdot (\mathcal{N}_{-3} + \mathcal{N}_{-1.5} + \mathcal{N}_{2.5} + \mathcal{N}_4 + \mathcal{N}_{11})$	0	Multi	0	Multi
New examples introduced in this chapter				
$0.9 \cdot \mathcal{N}(0,1) + 0.1 \cdot \mathcal{N}(5,1)$	1	Uni	0.6	Uni
$0.85 \cdot \mathcal{N}(0,1) + 0.15 \cdot \mathcal{U}([3,7])$	0.39	Uni	0.6	Uni

Table 1: Simulation results. P-values of the DFTU, and the dip test. $\mathcal{N}_a = \mathcal{N}(a, 0.5)$. Uni and Multi stand respectively for unimodal and multimodal.



(a) Histogram (b) ECDF

Figure 5: Histogram and empirical cumulative distribution function (ECDF) for the mixture $0.9 \cdot \mathcal{N}(0,1) + 0.1 \cdot \mathcal{N}(5,1)$.



(a) Histogram (b) ECDF

Figure 6: Histogram and empirical cumulative distribution function (ECDF) for the mixture $0.85 \cdot \mathcal{N}(0,1) + 0.15 \cdot \mathcal{U}([3,7])$.

References

A. Alfons, A. Archimbaud, K. Nordhausen, and A. Ruiz-Gazen. Tandem clustering with invariant coordinate selection. *Econometrics and Statistics*, 2024.

E. Anderson. The species problem in iris. *Annals of the Missouri Botanical Garden*, 23(3):457–509, 1936.

P. Arabie and L. J. Hubert. Cluster Analysis in Marketing Research. In R. P. Bagozzi, editor, *Advanced methods in marketing research*, pages 160–189. Blackwell, Oxford, 1994.

A. Archimbaud, K. Nordhausen, and A. Ruiz-Gazen. ICS for multivariate outlier detection with application to quality control. *Computational Statistics & Data Analysis*, 128:184–199, 2018.

C. Becquart. ICSpyLab: A Python package for invariant coordinate selection. *arXiv preprint*, 2026.

C. Becquart and A. Abdelsameia. ICSpyLab, 2026. Version 1.0.1.

C. Becquart, A. Archimbaud, A. M. Ruiz, and Z. Smida. A note on the folding test of unimodality: limitation and improved alternative. *arXiv preprint*, 2026.

N. Campbell and R. Mahon. A multivariate study of variation in two species of rock crab of the genus leptograpsus. *Australian Journal of Zoology*, 22(3):417–425, 1974.

R. B. D'Agostino, A. Belanger, and R. B. D'Agostino Jr. A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4):316–321, 1990.

R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

J. H. Friedman. Exploratory projection pursuit. *Journal of the American statistical association*, 82(397):249–266, 1987.

J. A. Hartigan and P. M. Hartigan. The dip test of unimodality. *The annals of Statistics*, pages 70–84, 1985.

H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933.

P. J. Huber. Projection pursuit. *The annals of Statistics*, pages 435–475, 1985.

I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, New York, 2 edition, 2002.

P. J. Rousseeuw and K. V. Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223, 1999.

A. Siffer, P.-A. Fouque, A. Termier, and C. Largouët. Are your data gathered? In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2210–2218, 2018.

B. W. Silverman. Using Kernel Density Estimates to Investigate Multimodality. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 43(1):97–99, 1981.

D. E. Tyler, F. Critchley, L. Dümbgen, and H. Oja. Invariant co-ordinate selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):549–592, 2009.