An Empirical Analysis of Topological Persistence as a Supplementary Measure of Dataset Drift

Christopher Shultz* and Simon Rudkin+

c.shultz@live.com

Author Information: *: c.shultz@live.com, Independent. +: simon.rudkin@manchester.ac.uk, Senior Lecturer in Data Science. University of Manchester. Equal contribution.

Keywords: dataset drift, covariate shift, dataset shift, topological data analysis, TDA

MSC: 55N31, 62-08

Abstract

This paper proposes the use of Persistence Entropy (PE), a concept from Topological Data Analysis (TDA) as a supplementary method of drift detection, particularly for its ability to capture higher-dimensional relationships that are unobservable to traditional drift detection mechanisms. We present a brief review of the drift detection literature and then discuss the benefits of using PE to capture the stability of the “shape” of data through time. Via numerical experiment, we demonstrate that the proposed method outperforms traditional drift detection strategies *and* is sensitive to higher-dimensional changes in the underlying structure of the data, capturing shifts that are missed by traditional tools.

1: Introduction

The concept of “drift” refers to systematic changes in a dataset over time, which may adversely impact model performance, affecting both the accuracy and fairness of models (Deho et al., 2024). Though often spoken about *generally*, drift can be categorized into *covariate shift*, in which the data distribution X evolves over time, but the relationship $X \sim Y$ remains stable; in *concept shift*, the relationship $X \sim Y$ changes over time (Mallick et al., 2022). A third type of drift in which some combination of covariate shift and concept shift take place simultaneously is also possible, and we refer to that case as *mixed shift.*

While unaddressed drift can lead to decaying model accuracy and fairness, several methods exist to detect it, each with its own set of advantages and limitations. Common approaches include the *Kolmogorov-Smirnov (KS)* test (compares distributions of two samples); *Hotelling's* T^2 test (compares a vector of means for a multidimensional dataset); and the *Population Stability Index (PSI)* (compares percentage membership within discrete bins for a single variable). Although they are straightforward, and serve similar, but distinct purposes, all fail to capture complex multivariable relationships.

In this paper, we propose the use of Topological Data Analysis (TDA) as a supplementary approach for drift detection. Specifically, we explore the quantified *Persistence Entropy (PE)* to assess the “stability” of a dataset’s shape and structure through time. TDA allows for a deep examination of geometric and topological features, yielding insights not captured by traditional methods. For example, one can use TDA to conceptualize a dataset as a noisy sample from some multidimensional object, the shape and topological characteristics of which we can examine, such as connected components, holes, and voids.

PE is a method derived from persistent homology that quantifies the complexity and variability of a dataset’s topological characteristics by first encapsulating information about the birth and death of topological features across different levels of *resolution* in a persistence diagram, then examining the “stability” of these persistence diagrams through time. In theory, datasets that remain topologically stable through time will have a stable vector of PE values. Datasets which experience changes in topological structure should accordingly have a PE vector that demonstrates jumps or drops that correspond to the change. By monitoring changes in PE over time, we can detect structural alterations in the data that indicate drift. This method is particularly valuable as supplement to traditional methods, as it identifies higher-order interactions and dependencies that are otherwise missed.

To validate this approach, we conduct an empirical study using simulated data. First, we generate a synthetic dataset of variables that follow a standard normal distribution, and then inject various types (and magnitudes) of drift toward the end of the series. We also compare the visual evolution of the quantified PE across these various scenarios and compare the performance of PE used as a drift detection measure against other traditional metrics.

Our findings demonstrate that [a] PE consistently captures the introduced drift through notable changes in computed average PE values; and [b] the magnitude of those changes aligns with the magnitude of applied shocks. Separately, PE outperforms traditional measures by capturing higher-dimensional information and changes in topological structure that are otherwise unobservable. By utilizing PE for drift detection, this work provides a novel approach that leverages topological information to provide a more comprehensive toolkit for maintaining the reliability and performance of predictive models.

The remainder of this paper is structured as follows. Section 2 provides an overview of drift, detection methods, and a background on TDA. Section 3 outlines the theoretical foundations of PE and its application in drift detection, along with the setup of an empirical study. Section 4 presents results and a discussion of findings. Finally, Section 5 concludes with a summary and some potential directions for future research.

2: Background

2.1: Overview of the Concept of Drift

The most basic form of drift is called *covariate shift*, which occurs when data is generated through some framework $P(Y|X)P(X)$ and the distribution $P(X)$ undergoes changes through time. This is referred to as covariate shift becaes only the covariate distribution changes (Quinonero-Candela et al., 2009). Described another way, this shift occurs when the mapping from $X \sim Y$ remains consistent, but the distribution of inputs (X) changes (Chen et al., 2016). The precise definition of covariate shift has been subject to debate, and aiming to unify terminology, Moreno-Torres et al. (2011) settle on a definition that aligns with the aforementioned explanation.

Another commonly discussed form of drift is *concept shift*, in which the relationship mapping $X \rightarrow Y$ changes from its prior form. For example, imagine a scenario where some model is built to predict Y using X, and the relationship underlying this structure is $Y = X^2$. If the relationship slowly changes over time to $Y = X^3$, this is an example of concept shift. A third type of drift, *mixed shift*, occurs when some combination of covariate shift and concept shift occur simultaneously.

All forms of drift have the potential to cause problems and are prevalent in real-world applications. Theoretical justification for most models relies on an assumed absence of drift (Tripuraneni et al., 2021), and with its presence, core assumptions are violated, meaning researchers and practitioners must understand how to detect it and its impacts on model validity, performance, and fairness.

The impacts of drift have been discussed in the literature for decades, both in terms of discussing the problem itself, as well as methods for its resolution. The most common problem we aim to consider is the degradation of model performance metrics like accuracy, precision, and recall. Widmer and Kubat (1996) discuss drift using the language of "changing context" and describe a family of algorithms that can flexibly react to a drifting environment through "online learning." This is one means by which attempts have been made to "correct for" the problems presented by drift.

The performance of classifiers, for example, have shown drastic improvement when covariate shift is detected and corrected for (Dharani et al., 2019). Failing to correct for degraded performance resulting from drift can lead to poor decision making, financial losses, and loss of trust in models by stakeholders. Gama et al. (2014) provide a comprehensive overview of the impacts of concept drift and covariate shift on model performance and methods for their detection.

Rather than another review of drift types, detection methods, and handling, we provide a brief overview resting on foundations set forth in other papers. We discuss the pros and cons of "standard" drift detection mechanisms and then present TDA as a supplementary measure capable of capturing otherwise unobservable information in the dynamics of the evolving dataset.

2.2: Traditional Methods for Drift Detection

Commonly used measures include the Kolmogorov-Smirnov (KS) test, the Hotelling T^2 test, and the Population Stability Index (PSI). Each has its own set of pros and cons, and all fail to capture the higher-dimensional relationships that TDA can provide, though we view the use of TDA as a supplement to these core methods, rather than a replacement.

2.2.1: Kolmogorov-Smirnov Test

The *Kolmogorov-Smirnov (KS)* test is a nonparametric statistical method for comparing the cumulative distribution functions (CDFs) of two univariate datasets. For each value therein, the KS test calculates the difference between CDFs, with the test statistic D representing the maximum absolute difference observed:

$$D = \sup_x |F_1(x) - F_2(x)|$$

where $F_1(x)$ and $F_2(x)$ represent the empirical CDFs of the two datasets, respectively. The null hypothesis H_0 states that the two datasets come from the same underlying distribution. If D is larger than the critical value (determined by the sample size and selected significance level), H_0 is rejected, indicating evidence for a significant difference between the two distributions (Massey, 1951; Smirnov, 1948).

Because the KS test is univariate, it cannot be applied simultaneously to all variables in a multidimensional dataset, and most applications will involve comparing the "old" and "new" series for each variable in isolation. For a truly multidimensional representation of drift, we may consider *Hotelling's* T^2 test, which we describe in the following section.

2.2.2: Hotelling's T^2 Test

Hotelling's T^2 test (Hotelling, 1931) is a multivariate statistical test that compares the means of multiple variables simultaneously, creating a more detailed window into the evolution over time of the joint distribution underlying the data. It is a generalization of the Student's t-test to the multivariate setting, and its test statistic is defined as:

$$T^2 = n(\bar{X} - \mu)^T S^{-1} (\bar{X} - \mu)$$

where n is the sample size, $\bar{X}$ is the sample mean vector, μ is the population mean vector, and S is the sample covariance matrix. The null hypothesis H_0 states that the sample mean vector is equivalent to the population mean vector, indicating equality of means. The test statistic follows a distribution related to the F-distribution, wherein if the calculated T^2 value exceeds the critical value from the F-distribution, the null is rejected, indicating a significant difference between mean vectors (Li et al., 2020). This allows for a multidimensional comparison of means, but still fails to capture higher-dimensional relationships and properties.

2.2.3: The Population Stability Index (PSI)

Perhaps most commonly used in practice, the *Population Stability Index (PSI)* measures the amount of change in a dataset based on a single variable, quantifying the change of the fraction of entities therein at several possible values/ranges (Haas and Sibbald, 2024). The PSI is usually stated as:

$$PSI = \sum_{bins,i} \left(f_{1,i} - f_{0,i}\right) \cdot \ln\left(\frac{f_{1,i}}{f_{0,i}}\right)$$

with $\left(f_{0,i}, f_{1,i}\right)$ representing the fraction of entities in bin i in the original and new populations, respectively. As a result, this method necessitates the binning of each continuous variable into discrete groups before application. The PSI is univariate in nature and results in the comparison of individual variables in isolation, though their results can be examined in aggregate, similar to the KS test.

Kurian and Allali (2024) frame PSI as a variant of KL divergence, wherein given two probability distributions the KL divergence measures the "excess surprise" in using the actual distribution vs the expected distribution. They point out the core limitation that KL divergence is not symmetric (i.e. given two datasets Q, P, the $D_{KL}(Q||P) \neq D_{KL}(P||Q)$), whereas PSI resolves this problem by modifying KL divergence into a symmetric measure.

The PSI does provide a more granular view of the stability of the underlying distributions than the simple mean method using in Hotelling's T^2 test, but still fails to capture higher-dimensional interactions and "shape" features such as joint distribution changes and topological features such as holes or voids. Yurdakul and Naranjo (2020) provide a thorough overview of the statistical properties of the PSI and a summary of its pros and cons.

2.2.4: Summary of Standard Drift Detection Methods

As we have seen, most commonly used drift detection methods are powerful, but have limited scope. Traditional univariate methods such as the KS test or PSI fail to examine the joint distribution and evolving "shape" of the combined dataset, focusing rather on the changing distributions of individual variables alone. Multivariate tools like Hotelling's T^2 provide a "combined" view of the representation of data, but are limited by a focus only on comparison of means vectors, rather than distributions at large. For a more complete view, Rabanser et al. (2019) provide a thorough review of statistical methods, outlining their empirical structure, benefits, and weaknesses.

2.3: Introducing Topological Data Analysis (TDA)

The stated limitations of traditional drift detection methods are related to their scope, not their validity. The proposal herein is that methods from TDA can *supplement* traditional methods by providing a new window into the dynamic shape and topological characteristics of data, illuminating information hat is otherwise unobservable. The use of standard drift detection tools has been widely adopted, but fails to paint the entire picture of what's happening as datasets evolve and change through time.

Viewing a dataset as a noisy sample from some multidimensional "object" with shape and form, methods from TDA provide a framework for capturing information about the underlying structure and interrelationships embedded in data, further informing our concept of "drift." In brief, the question we aim to answer is this: *does the shape and topological structure of a given multidimensional dataset change over time, and can that change be quantified in a way that illuminates drift?*

The general framework employed in many applications of TDA is to first represent a set of data as a *point cloud*, this point cloud as a *simplicial complex* defined by some resolution parameter, and a collection of these simplicial complexes across varying levels of resolution as a *filtration*. This TDA pipeline is summarized by Lum et al. (2013).

More formally, the original dataset forms a set of discrete points in some space (commonly Euclidean), which we refer to as a point cloud. This point cloud is viewed as a noisy sampling of observations from some underlying structure which can then be estimated (Carlsson, 2009). To accomplish this, we construct a *simplicial complex*, which is a "smoothed" representation of the point cloud, constructed as the vertices of a combinatorial graph whose edges are determined by a proximity measure (ϵ) defining the "resolution" of the complex (Lum et al., 2013). A common choice in deriving the simplicial complex is the Vietoris-Rips algorithm, summarized by Shultz (2023).

The development of multiple simplicial complexes at various levels of scaling (ϵ) creates a *filtration* of complexes $\emptyset \subseteq \mathbb{X}_0 \subseteq \mathbb{X}_1 \subseteq \cdots \subseteq \mathbb{X}_n \subseteq \mathbb{X}$. In this context, we consider the "birth" and "death" values of the topological features that emerge (e.g. connected components, holes, voids), and the values of resolution through which they persist (Gidea and Katz, 2018). *Persistence Diagrams* encode this information visually to display birth-death pairs for each observed topological feature within the filtration, with the idea that "longer-lasting" features are significant, whereas those which are born and quickly die are likely due to noise (Zomorodian and Carlsson, 2005).

For the detection of drift in an evolving dataset, we propose the use of *Persistence Entropy (PE),* which generates a vector of quantified entropy representing the amount of order/disorder in the underlying topological object through time (Munch et al., 2019). For example, with a series of data X over some time index t, we can utilize a rolling-window approach with a user-selected window length (e.g. $w = 30$) to create sequential sub-frames, each of length w. We then apply this rolling window to the TDA pipeline, each first represented as a point cloud, converted to simplicial complexes, then to persistence diagrams which are used in the PE computation.

Given a persistence diagram of birth-death-dimension triples (b, d, q), persistence entropies are calculated as the base-2 Shannon entropies of the collections of differences $(d - b)$ "lifetimes", normalized by the sum of all such differences. Formally, the PE is represented as $H(X)$ below.

$$H(X) = -\sum_i \left(\frac{l_i}{L}\right) \log \left(\frac{l_i}{L}\right)$$

where l_i/L represents the normalized persistence of each bar in the diagram, with l_i being the length of the i_{th} bar and L the sum of the lengths of all bars in the diagram. This quantifies the "complexity" of the persistence diagrams via the concept of entropy. The value stream $H(X)$ creates a window into a higher level of dimensionality than traditional drift detection methods allow for; and tracking this measurement can provide a view into the stability of the topological complexity of the underlying object, which was previously invisible. For example, it is possible that means remain stable, but topological structure drifts, and this would be missed in traditional approaches. In fact, this very outcome takes in our experiment described in Section 4. For a more thorough introduction to TDA, refer to Chazal and Michel (2017).

To our knowledge, no existing research examines the utilization of TDA and/or PE computations for the detection of dataset drift. The utilization of TDA-based methods for comparing distributions, however, is an emerging area of research, with Dlotko et al. (2024) examining a comparison of distributions topologically using the Euler Characteristic Curve (ECC). Their analysis shows that TDA-based comparisons perform similarly to state-of-the-art methods, lending further support for the exploration of topology-based comparisons in drift detection.

3: Methodology

To explore the question of whether PE can capture drift, we simulate a dataset for experimentation. The simulated data $x_1, x_2, x_3 \in X$ consists of three standard normal variates $x_i \sim N(0,1)$ of length $N = 1000$. We derive a fourth variable y as a linear combination of the three covariates such that $y = \sum_i x_i$. Together, this creates a baseline dataset $D_{1000\times4}$. We then intentionally inject various forms (and magnitudes) of drift into the system with a series of modifier tables of size $M_{400\times4}$, each of which is appended to the bottom of D, forming a set of final datasets $\mathcal{D}_{s_{1400\times4}}$, where $s \in S$ represents a series of "drift scenarios."

First, we simulate covariate shift by altering the distribution of x_1 via changes to its variance such that $x_1 \in M_{400\times4}$ is distributed $N(0, v)$, where v is a parameter in the set {0.0001, 0.001, 0.01, 0.1, ,0.6, 1, 5, 10, 20, 30}. This allows for the examination of the impact of changing variance alone *and* at different magnitudes. Note that in the case of $v = 1$, this is the same as "no drift", as the modifier dataset appended is distributed identically to the baseline.

Second, we introduce concept shift by altering the relationship between X, Y in the modifier system after time T = 1000. In the original setting, $D_{1000\times4}$ has $y = x_1 + x_2 + x_3$. In the modifier setting, the dataset M has $y = x_1^v + x_2 + x_3$, where v is in the set $\{2, 3, 4, 0.5, 0.25\}$. We also consider the modified linear case $y = 2x_1 + x_2 + x_3$ under this umbrella.

Finally, we consider the examination of "mixed" drift, which combines both covariate shift and concept shift simultaneously, across various combinations of the covariate shift parameter applied to the case when y is changed to the relationship $y = x_1^{0.5} + x_2 + x_3$.

As a result, we have a number of "modified" datasets, which have the same base $D_{1000\times4}$ appended with various forms of $M_{400\times4}$, representing the various forms of drift discussed, leading to multiple instances of $\mathcal{D}_{s_{1400\times4}}$, each $s \in S$ representing a different drift scenario. Our objective is to consider two questions: [a] how does the quantified persistence entropy vector change visually across time, when drift is introduced in various forms/magnitudes?; and [b] how do various forms of drift detection compare to the proposed PE method?

To examine question [a], we consider each dataset $D_{s_{1400\times4}}$ individually, loop through the time index via a rolling window ($w = 30$) such that the first window consists of points $t \in [0{:}\,29]$, the second window consists of points $t \in [1{:}\,30]$, and so on. Each window creates a subset $W_{30\times4} \subset \mathcal{D}_{s_{1400\times4}}$ where W constitutes a "point cloud" of a snapshot in time. We construct a Vietoris-Rips simplicial complex on top of this point cloud, and build a persistence diagram to consider the birth-death times of the observed topological features. We then use PE to quantify the entropy of persistence diagrams over the sliding time horizon. Intuitively, if the topological structure underlying the data changes significantly, we should expect an observable change to PE around the same time. We provide a number of visualizations to plot the course of the PE vector from the beginning of the data series to the end, showing that the PE vector jumps or drops when drift is introduced, typically scaling with the magnitude of drift selected.

To examine question [b], we take a single case of strong drift, and examine the performance of various drift detection methods in capturing the drift injected. We utilize the special case where the $M_{400\times4}$ appended to the baseline dataset is modified with covariate shift $x_1 \sim N(0,30)$ and concept shift in that the relationship $X \sim Y$ is changed to $y = x_1^{0.5} + x_2 + x_3$. We then use a rolling window approach to create various "before" and "after" datasets, which we use as inputs to our drift detection tests. The selected rolling window is of length 200, so the "before" set is $B_{T-200:T}$ and the "after" set is $A_{T+1:T+201}$. In Section 4, we describe the results of our experiments on the simulated data, and discuss intuition and implications for each outcome.

4: Results

Table 1 below presents a summary of the simulated datasets $D_{s_{1400\times4}}$ used in our analysis, wherein the first 1000 observations are identically distributed, and the remaining 400 observations are drifted in various ways and magnitudes depending on the specific scenario s. We present the summary statistics for x_1 and y in particular, because x_2 and x_3 remain identically distributed across all scenarios. It is apparent that variance scales with the magnitude of covariate shift injected, though means remain near zero.

Table 1: Summary Statistics

Drift Scenario		X1				Y			
Type	**Factor**	**min**	**max**	**mean**	**var**	**min**	**max**	**mean**	**var**
Covariate Shift	0.0001	-3.24	3.85	0.01	0.68	-4.86	5.6	0.05	2.65
Covariate Shift	0.01	-3.24	3.85	0.01	0.68	-4.56	5.6	0.05	2.66
Covariate Shift	1	-3.24	3.85	0.01	0.95	-4.62	5.6	0.05	2.79
Covariate Shift	10	-33.21	32.87	0.07	32.37	-32.89	31.2	0.12	33.42
Covariate Shift	30	-88.61	98.57	0.35	280.25	-87.64	98.58	0.43	282.03
Concept Shift	x^2	-3.54	3.85	-0.01	0.98	-4.56	11.64	0.37	3.47
Concept Shift	x^4	-3.24	3.85	0.03	0.99	-4.59	90.89	0.99	27.18
Concept Shift	x^0.5	-3.24	3.94	0	0.96	-4.59	5.6	0.21	2.84
Concept Shift	x^0.25	-3.33	3.85	0.01	1	-4.59	5.6	0.23	2.8
Mix Shift	Covariate Shift 30 Concept Shift 0.5	-68.15	111.84	0.87	263.7	-4.59	11.24	0.89	6.44

4.1: Question A: How Does Persistence Entropy Evolve Over Time With Drift?

For each of the individual drift types (covariate shift, concept shift, mixed) and magnitudes, we plot the average PE value over 50-length blocks (for easier visualization), observing that the PE time series demonstrates two interesting properties. First, the PE value either spikes or drops when drift is added to the dataset. This depends somewhat upon the type and "direction" of the drift, but movements are systematically observed. Second, the magnitude of the observed change in PE corresponds to the magnitude of the drift injected. This means that small changes in topological structure are likely to go unnoticed, whereas larger changes should be visibly obvious. It is also possible to apply formal rules to the evolution of PE for example, by conducting a t-test to estimate when the mean changes significantly between old and new datasets. We employ this exact approach to assess the performance of PE for drift detection to answer question [b].

Figure 1 below illustrates the evolution of the quantified PE vector through the time horizon groups (indexed in chunks of 50 on the x-axis). We can observe that covariate shift impacts the evolution of the time series significantly, except in the case of shift $v = 1$, which corresponds to the "no drift" case. For high values of v, i.e. $v > 1$, we observe a marked drop in the PE vector, the magnitude of which grows the further we move from 1. Where $v < 1$, we observe an upward shift in the quantified PE vector.

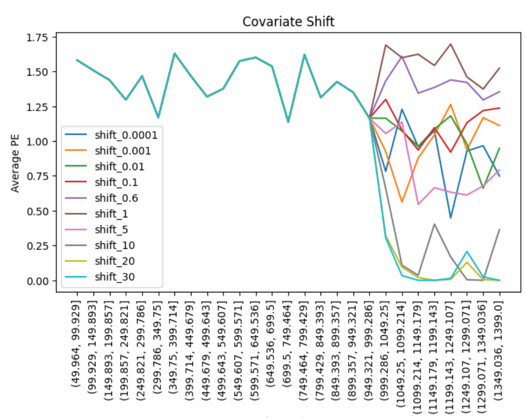

Figure 1: Evolution of PE with Covariate Shift

Figure 2 shows the impact of concept shift on the evolution of PE through time. Recall that the original relationship $y = x_1 + x_2 + x_3$ which holds from $T \in [0{:}\,1000]$ is modified to $y = x_1^v + x_2 + x_3$ where $T \in [1001{:}\,1400]$. Interestingly, the resulting PE vector does not change significantly when $v > 1$, but drops extremely fast and then stabilizes when $x < 1$. A similar relationship is visible in Figure 3, which illustrates mixed shift, wherein we have $x_1 \sim N(0, v)$ and $y = x_1^{0.5} + x_2 + x_3$. In this case, the concept shift introduced seems to dominate, and regardless of which parameter we select for v, a major drop occurs in PE around $T = 1000$.

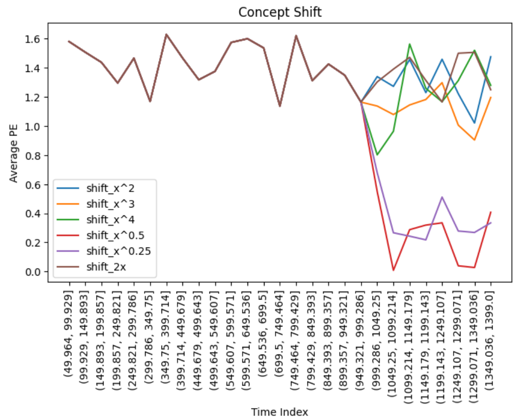

Figure 2: Evolution of PE with Concept Shift

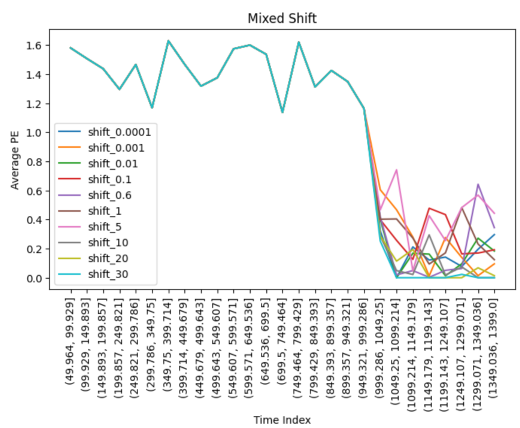

Figure 3: Evolution of PE with Mixed Shift

The collection of Figures 1-3 demonstrate that the PE method *is* capable of detecting drift of both the *covariate shift* and *concept shift* varieties. Though the magnitude of the change to the PE vector does depend on the strength and type of drift introduced. We next consider whether we can formalize a statistical procedure to use PE as a means of detecting dataset drift in a production setting, and comparing its performance to classical drift detection metrics.

4.2: Question B: How Does PE Perform As a Drift Detection Mechanism?

As described in Section 3, we utilize the special case of strong drift where the M_{400x4} set is modified with $x_1 \sim N(0{,}30)$ and $y = x_1^{0.5} + x_2 + x_3$. We then utilize a rolling window approach to create "before" and "after" datasets, which are used to assess whether drift is identified between the two subdivisions of the data stream. This rolling window is of length 200, so the "before" set is $B_{T-200:T}$ and the "after" set is

$A_{T+1:T+201}$. We then compare the ability of PE to detect drift on the combined dataset $\mathcal{D}_{1400\times 4}$ given that we know drift enters the system at T = 1000. To formalize the "test" process for using PE, we consider the vector of PE values as a stream on which we want to consider the stability of the mean. The premise is that if the mean of PE remains stable from $B \to A$, then drift is unlikely, whereas if the mean jumps or drops significantly, drift is likely. To compare this, we utilize a t-test. Table 2 provides an overview of the tests performed in our comparison.

Table 2: Overview of Tests for Drift

Test	**Type**	**Notes**
Kolmogorov-Smirnov (KS)	Univariate	Considers only x_1. Drift indicated if p-value < 0.001.
Hotelling's T^2 (H2)	Multivariate	Considers the full set (x_1, x_2, x_3, y). Drift indicated if p-value < 0.001.
Population Stability Index (PSI)	Univariate	Considers only x_1. Drift indicated if PSI statistic > 0.25.
Persistence Entropy (PE)	Multivariate	Considers the full set (x_1, x_2, x_3, y). Computes PE vector and uses t-test to compare the means from the old and new sets. Drift indicated if p-value < 0.001.

The results of this comparison are provided in Figure 4 below. In general, the KS test performs well at capturing drift injected into x_1, though with a considerable delay beyond the time of injection. The PSI similarly performs well, though with some false positives in the time period T < 1000, and a drop-off corresponding to the size of the rolling window that defines the "before" and "after" datasets. From a multivariate perspective, the H2 test does a poor job of identifying the drift injected, though when it does indicate drift, those cases are all true positives. Despite some false positive cases when T < 1000, the PE test strongly identifies drift within the system.

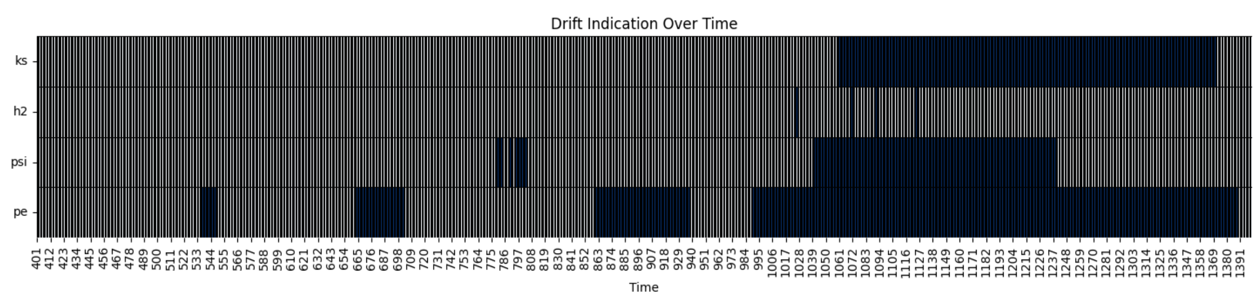

Figure 4: Comparison of Test Performance

Considering H2 and PE, both multivariate tests, we refer back to their underlying intuition to explain the stark difference in outcomes. H2 almost completely fails to capture the drift injected, whereas PE captures it excellently, if not a bit too excitedely, resulting in some false positives. This occurs because while H2 is multivariate in nature, it only considers the comparison of means vectors, whereas the PE method captures complex high-dimensional structure and dependencies underyling the dataset.

This illustrates the situation in which traditional multivariate metrics such as H2 fail to alert the user of drift that is quickly and readily identified by PE. Recall in particular that the injected drift does not modify the means underlying the system, and as a result, the significant changes to the topological structure of the data are missed by the H2 test due to its reliance on means vector comparisons. While examining the univariate metrics such as KS and PSI in isolation also performs well, neither capture the whole picture, and PE demonstrates the ability to capture complex interdependence changes that are otherwise invisible.

In summary the PE method shows excellent performance in identifying injected structural changes, despite some brief false positives. This may be particularly powerful in a field like finance, whereas commonly modeled values such as asset returns traditionally have mean zero.

5: Conclusion

In this paper, we introduced Persistence Entropy (PE), a measure from Topological Data Analysis (TDA), formalized via t-test, as a supplementary method for detecting dataset drift. Through empirical analysis, we demonstrate that PE effectively captures introduced drift, showing consistent changes corresponding to the magnitude of the applied drift. We also show that it outperforms traditional measures of drift detection, particularly in its ability to capture complex, higher-dimensional relationships between variables, exposing the underlying topological structure.

These results support the hypothesis that PE can detect subtle and complex topological changes that are overlooked by conventional methods, meaning that its use could prevent adverse outcomes that might otherwise be missed. By computing and tracking PE over time, practitioners can gain deeper insights into the evolving data landscape, enhancing their ability to maintain model performance and reliability.

Integrating TDA into existing drift detection frameworks provides a more comprehensive toolkit for addressing drift, particularly in areas like finance, where mean-zero vectors are common (e.g. asset returns). Directions for future research include the application of PE in various real-world scenarios, considering its scalability to larger datasets and the computational complexity problem, and its sensitivity to the selection of various parameters such as sliding window size in the persistence entropy calculation. Overall, this study contributes to the growing body of literature on drift detection, offering a novel approach to leverage topological information to supplement traditional methods.

References

Carlsson, G. 2009. Topology and Data. *Bulletin of the American Mathematical Society.* 46(2): 255-308.

Chazal, F. and B. Michel. 2017. An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. *arXiv: 1710.04019.*

Chen, X., Monfort, M., Liu, A., and B. Ziebart. 2016. Robust Covariate Shift Regression. *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS) 2016.* Cadiz, Spain.

Deho, O., Liu, L., Li, J., Liu, J., Zhan, C., and S. Joksmiovic. 2024. When the Past != The Future: Assessing the Impact of Dataset Drift on the Fairness of Learning Analytics Models. Accepted in *IEEE Transactions on Learning Technologies.*

Dharani, G., Nair, N., Satpathy, P., Christopher, J. 2019. Covariate Shift: A Review and Analysis on Classifiers. *2019 Global Conference for Advancement in Technology. Bangalore, India.*

Dlotko, P., Hellmer, N., Stettner, L. and R. Topolnicki. 2024. Topology-Driven Goodness-of-Fit Tests in Arbitrary Dimensions. *Statistics and Computing.* 34:34.

Gama, J., Zliobaite, I., Bifet, A., Pechenizkiy, M., and A. Bouchachia. 2014. A Survey on Concept Drift Adaptation. *ACM Computing Surveys (CSUR).* 46(4): 1-37.

Gidea, M. and Y. Katz. 2018. Topological Data Analysis of Financial Time Series: Landscapes of Crashes. *Physica A: Statistical Mechanics and Its Applications.* 491.

Haas, M. and L. Sibbald. 2024. Measuring Data Drift with the Unstable Population Indicator. *Data Science.* 7: 1-12.

Hotelling, H. 1931. The Generalization of Student's Ratio. *Annals of Mathematical Statistics.* 2(3): 360-378.

Huang, Y., Yuan, Z., Leung, C., Wu, Q., Ma, S., Wang, S., Wang, D., and Z. Huang. 2023. Towards Balanced Representation Learning for Credit Policy Evaluation. *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS) 2023.* Valencia, Spain.

Kurian, J. and M. Allali. 2024. Detecting Drifts in Data Streams Using Kullback-Leibler (KL) Divergence Measure for Data Engineering Applications. *Journal of Data, Information, and Management.*

Li, H., Aue, A., Paul, D., Peng, J. and P. Wang. 2020. An Adaptable Generalization of Hotelling's T^2 Test in High Dimension. *The Annals of Statistics.* 48(3): 1815-1847.

Lum, P., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J. and G. Carlsson. 2013. Extracting Insights from the Shape of Complex Data Using Topology. *Scientific Reports.* 3(1): 1-8.

Mallick, A., Hsieh, K., Behnaz, A., and G. Joshi. 2022. Matchmaker: Data Drift Mitigation in Machine Learning for Large-Scale Systems. *Proceedings of the 5th MLSys Conference,* Santa Clara, CA, USA.

Massey, F. 1951. The Kolmogorov-Smirnov Test for Goodness of Fit. *The Journal of the American Statistical Association.* 46(253): 68-78.

Moreno-Torres, J., Raeder, T., Alaiz-Rodriguez, R., Chawla, N., and F. Herrera. 2011. A Unifying View on Dataset Shift in Classification. *Pattern Recognition.* 45: 521-530.

Munch, E., Myers, A., and F. Khasawneh. 2019. Persistent Homology of Complex Networks for Dynamic State Detection. *Physical Review E.* 100.

Quinonero-Candela, J., Sugiyama, M., Schwaighofer, A., and N.D. Lawrence. 2009. *Dataset Shift in Machine Learning.* The MIT Press.

Rabanser, S., Gunnemann, S., and Z. Lipton. 2019. Failing Loudly: An Empirical Study of Methods for Detecting Dataset Shift. *33rd Conference on Neural Information Processing Systems.* Vancouver, CAN.

Shultz, C. 2023. Applications of Topological Data Analysis in Economics. *SSRN.* https://ssrn.com/abstract=4378151.

Smirnov, N. 1948. Table for Estimating the Goodness of Fit of Empirical Distributions. *Annals of Mathematical Statistics.* 19(2). 279-281.

Tripuraneni, N., Adlam, B., and J. Pennington. 2021. Overparameterization Improves Robustness to Covariate Shift in High Dimensions. *35th Conference on Neural Information Processing Systems.*

Widmer, G. and M. Kubat. 1996. Learning in the Presence of Concept Drift and Hidden Contexts. *Machine Learning.* 23(1): 69-101.

Yurdakul, B. and J. Naranjo. 2020. Statistical Properties of the Population Stability Index. *Journal of Risk Model Validation.* 14(4): 89-100.

Zomorodian, A. and G. Carlsson. 2005. Computing Persistent Homology. *Discrete & Computational Geometry.* 33(2): 249-274.