STATISTICAL PERSISTENT HOMOLOGY OF BRAIN SIGNALS

Yuan Wang[1]*, Hernando Ombao*[2]*, Moo K. Chung*[3]

[1]University of South Carolina, U.S.A.
[2]King Abdullah University of Science and Technology, Saudi Arabia
[3]University of Wisconsin-Madison, U.S.A.

ABSTRACT

Topological data analysis (TDA) extracts hidden topological features in signals that cannot be easily decoded by standard signal processing tools. A key TDA method is persistent homology (PH), which summarizes the changes of connected components in a signal through a multiscale descriptor such as the persistent landscape (PL). A recent development indicates that statistical inference on PLs of scalp electroencephalographic (EEG) signals produces markers for localizing seizure foci. However, a key obstacle of applying PH to large-scale clinical EEGs is the ambiguity of performing statistical inference. To address this problem, we develop a unified permutation-based inference framework for testing statistical indifference in PLs of EEG signals before and during an epileptic seizure. Compared with the standard permutation test, the proposed framework is shown to have more robustness when signals undergo non-topological changes and more sensitivity when topological changes occur. Furthermore, the proposed new method drastically improves the average computation time by 15000 folds.

***Index Terms*—** EEG, persistent homology, persistence landscape, exact permutation test

1. INTRODUCTION

Topological data analysis (TDA) extracts topological and geometric features in complex data that cannot be easily decoded by standard analytic tools [1]. A key TDA algorithm is persistent homology (PH), which reveals the topological changes of connected components and holes in the data through a multiscale descriptor such as the persistent landscape (PL). Recent development indicates that statistical indifference in PLs of scalp electroencephalographic (EEG) signals before and during seizure could help localize seizure foci [2]. This is quantified through a permutation test where labels of each frequency in the signals are permuted under the null hypothesis of identical distributions of the PLs before and during seizure. The standard permutation test is numerically shown to be robust under non-topological changes while being sensitive to topological changes in the signals.

Despite the benefits of detecting true topological changes, the standard permutation test requires at least 10000 iterations averaging 5 hours to converge for a 5-minute EEG recording. Such computational bottleneck is the key obstacle for clinical implementation, where close to real-time signal detection is needed. To overcome this computational challenge, we propose an exact permutation test on the PLs of EEG signals. The test computes an exact p-value by combinatorially enumerating all the possible permutations between two monotone statistics based on areas under the layers of the PLs, thus avoiding computationally intensive resampling used in the standard permutation test.

Compared with the standard permutation test, the proposed framework is shown to have more robustness when signals undergo non-topological changes and more sensitivity when topological changes occur. Furthermore, the proposed new method drastically improves the average computation time by 15000 folds. To the best of our knowledge, this is the first work proposing an exact permutation test to compare PH features in EEG signals.

2. METHODS

2.1. Denoising with weighted Fourier series

We denoise an EEG signal $f(t)$ recorded at regular time intervals $-T = t_1 < t_1 < \cdots < t_N = T$ by estimating the underlying signal $\mu(t)$ in the model:

$$f(t) = \mu(t) + \epsilon(t), -T \leq t \leq T, \tag{1}$$

with the series solution to the linear diffusion equation

$$\frac{\partial}{\partial \sigma} g(t,\sigma) = \frac{\partial^2}{\partial t^2} g(t,\sigma), \sigma \geq 0, t \in [-T,T]. \tag{2}$$

By treating the observed signal $f(t)$ as the initial condition of the diffusion equation: $g(t,0) = f(t)$, we are able to obtain a unique closed-form solution to (2) - a weighted Fourier series (WFS) approximation for the signal $\mu(t)$:

$$\widehat{\mu}(t) = \sum_{j=0}^{\infty} e^{-\gamma_j \sigma} a_j \phi_{j1}(t) + \sum_{j=1}^{\infty} e^{-\gamma_j \sigma} b_j \phi_{j2}(t), \tag{3}$$

with the eigenvalues $\gamma_j = (j\pi/T)^2$, the Fourier coefficients

$$\begin{aligned} a_0 &= \frac{1}{2T}\int_{-T}^{T} f(t)dt, \\ a_j &= \frac{1}{T}\int_{-T}^{T} f(t)\cos(j\pi t/T)dt, \\ b_j &= \frac{1}{T}\int_{-T}^{T} f(t)\sin(j\pi t/T)dt, \end{aligned}$$

and the cosine and sine basis functions ϕ_{j1} and ϕ_{j2}, $j \geq 0$. The degree-k representation of (3) is

$$\widehat{\mu}^k(t) = a_0 + \sum_{j=1}^{k} e^{-(j\pi/T)^2\sigma}[a_j\cos(j\pi t/T) + b_j\sin(j\pi t/T)], \tag{4}$$

where the degree k determines the highest frequency $[k/T]$ to be included in the representation and bandwidth σ tunes smoothness by controlling the rate of the decay of the weights corresponding to higher frequencies. In this paper, we use $k = 499$ and $\sigma = 0.0005$ for $T = 5$ unless otherwise stated.

2.2. Persistence landscape

PH tracks the changes of connected components in the sublevel set of $\widehat{\mu}^k$, and a barcode is a collection of intervals with endpoints being the birth and death times of the connected components as the horizontal threshold λ increases (Figure 1 top). The PL is a PH feature proposed in [3] for the purpose of establishing a statistical framework on PH. The easiest way to construct a PL is through its corresponding barcode. Given a bar $(\lambda_{i1}, \lambda_{i2})$ in a barcode $\{(\lambda_{1i}, \lambda_{2i})\}_{i=1}^N$, we can define the piecewise linear bump function $h_{(\lambda_{1i},\lambda_{2i})} : \mathbb{R} \to \mathbb{R}$ by

$$h_{(\lambda_{1i},\lambda_{2i})}(\lambda) = \max(\min(\lambda - \lambda_{1i}, \lambda_{2i} - \lambda), 0). \tag{5}$$

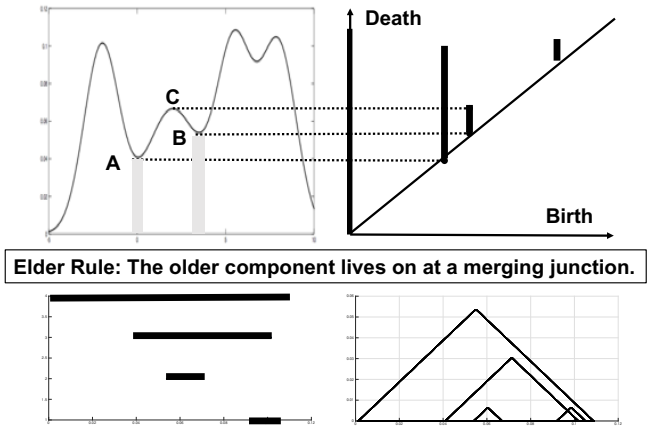

Fig. 1. Persistent homology, barcode and persistence landscape of a denoised EEG signal.

The PL ν of the barcode $\{(\lambda_{1i}, \lambda_{2i})\}_{i=1}^N$ is then defined as a multi-valued function with the ℓth value being

$$\nu_l(\lambda) = \begin{cases} l\text{-th largest value of } \{h_{(\lambda_{1i},\lambda_{2i})}(\lambda)\}_{i=1}^N & 1 \leq l \leq N, \\ 0 & l > N. \end{cases} \tag{6}$$

Figure 1 (bottom right) shows the PL of the barcode (bottom left). The main technical advantage of PL is that, as a random variable with values in a Banach space, it obeys a strong law of large numbers and a central limit theorem, in contrast to the barcode [3].

2.3. Exact permutation test on areas under PL layers

We propose the exact permutation test proposed in [4] to compare the PLs of two EEG signals: $\{\nu^1\}$ and $\{\nu^2\}$, e.g. before and during seizure. For each group $i = 1, 2$, we define the statistic A_ℓ^i as the value of the area under the PL layer ℓ, which is the reversely ordered layer number (Figure 2). Note that it reverses the defined order of PL layers in (6) from the most prominent to the least prominent layer.

The exact permutation test is then applied to A_ℓ^1 and A_ℓ^2, $\ell = 1, \ldots, L$, for comparison. The PL with a smaller number of layers is padded with zeros in the vectors of areas. Under the null hypothesis of equivalence of the two PLs, two PLs ν^1 and ν^2 are assumed to have identical distribution. Thus, we test the statistical difference between the areas

$$A_1^1, \ldots, A_L^1$$

and

$$A_1^2, \ldots, A_L^2$$

under the reversely ordered layers of ν^1 and ν^2:

$$H_0 : A_\ell^1 = A_\ell^2,$$

for all $\ell = 1, \ldots, L$.

As pointed out by [5], permutation test is not assumption-free. The assumption is that whatever feature or structure is permuted should be exchangeable under the null hypothesis. Here the proposed permutation test is built on the assumption that the layers in the PLs representing two groups are exchangeable. This is plausible because the null assumption is that there is no statistical difference between PLs before and during seizure and so the layers of the PLs do not differ significantly.

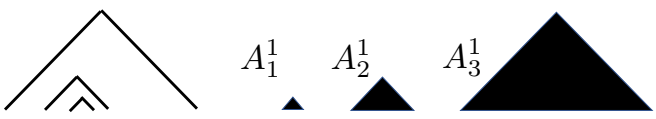

Fig. 2. Areas under the layers of a persistence landscape.

Note that $A_1^1 \leq \cdots \leq A_L^1$ and $A_1^2 \leq \cdots \leq A_L^2$. We define the corresponding monotone step functions $\psi_1(t)$ and $\psi_2(t)$:

$$\psi_1(t) = \begin{cases} 0 & \text{if } t < A_1^1, \\ \ell & \text{if } A_\ell^1 \leq t < A_{\ell+1}^1, 1 \leq \ell < L, \\ L & \text{if } t \geq A_L^1. \end{cases}$$

and

$$\psi_2(t) = \begin{cases} 0 & \text{if } t < A_1^2, \\ \ell & \text{if } A_\ell^2 \leq t < A_{\ell+1}^2, 1 \leq \ell < L, \\ L & \text{if } t \geq A_L^2. \end{cases}$$

Then we have the test statistic

$$D(\psi_1, \psi_2) = \sup_t |\psi_1(t) - \psi_2(t)|.$$

The *exact* p-value of an observed value d of the test statistic D is combinatorially given by

$$P(D \geq d) = 1 - \frac{A(L, L)}{C(2L, L)},$$

where $A(L, L)$ is the iteratively computed as

$$A(u, v) = A(u - 1, v) + A(u, v - 1)$$

with the boundary condition

$$A(0, L) = A(L, 0) = 1$$

within domain $|u - v| < d$, and

$$C(2L, L) = \frac{(2L)!}{L!L!}$$

is the total number of combinations of choosing L out of $2L$ [4]. Figure 3 displays a schematic for the exact permutation test on 3-layer PLs.

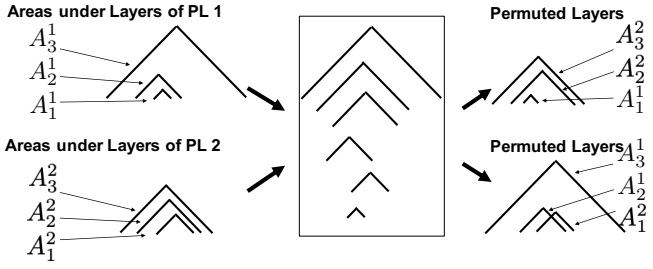

Fig. 3. Schematic of the exact permutation test on 3-layer persistence landscapes.

3. SIMULATIONS

We present two simulation studies to show the robustness and sensitivity of the exact test on PLs, compared with the standard permutation test in [2] and the paired t-test on local variance.

3.1. Study 1: Testing robustness.

This study shows the robustness of the proposed test under frequency scaling, a non-topological change, of the underlying signals. We simulate four signals $y(t_i)$ at regular time intervals $0 \leq t_1, \ldots, t_{500} \leq 2\pi$:

$$y(t_i) = t_i \cos(\omega t_i), \tag{7}$$

where ω takes on four values: 1) $\omega = 1$; 2) $\omega = 2$; 3) $\omega = 5$; 4) $\omega = 10$. In each simulation, independent Gaussian noises $N(0, 2^2)$ are added to the signals at $0 \leq t_1, \ldots, t_{500} \leq 2\pi$:

$$y(t_i) = t_i \cos(\omega t_i) + \epsilon_i,$$

where $\epsilon_i \sim N(0, 2^2)$. Results in Table 1 show that the paired t-test on local variance is sensitive to frequency scaling, whereas the permutation tests are robust to the non-topological changes at the 5% significance level.

3.2. Study 2: Testing sensitivity.

This study shows the sensitivity of the proposed test to certain topological changes in the underlying signals. We simulate four pairs of signals $y_1(t_i)$ and $y_2(t_i)$, at regular time intervals $0 \leq t_1, \ldots, t_{500} \leq 2\pi$:

$$y_1(t_i) = t \cos(\omega t_i),$$

$$y_2(t_i) = \begin{cases} y_1(t_i) & 0 \leq t_i \leq 0.4\pi, \\ y_1(t_i) - 200 & 0.4\pi < t_i \leq 0.96\pi, \\ y_1(t_i) + 200 & 0.96\pi < t_i \leq 1.04\pi, \\ y_1(t_i) - 200 & 1.04\pi < t_i \leq 1, 6\pi, \\ y_1(t_i) & 1, 6\pi < t_i \leq 2\pi, \end{cases}$$

where ω takes on one of four values: 1) $\omega = 1$; 2) $\omega = 2$; 3) $\omega = 5$; 4) $\omega = 10$. In each simulation, independent Gaussian noise $N(0, 50^2)$ are added to the signals at $0 \leq t_1, \ldots, t_{500} \leq 2\pi$. Since the discontinuities in y_2 introduce topological tearing to the signals, we test weather y_1 and y_2 have different PLs. Results in Table 1 show that all three tests are fairly sensitive to the topological tearing in the signals at the 5% significance level.

Study 1	$\omega = 1$ vs 2	$\omega = 1$ vs 5	$\omega = 1$ vs 10
Variance	14.7%	79.8%	99.9%
Standard	0%	0%	0.1%
Exact	0%	0%	0%
Study 2	$\omega = 1$	$\omega = 5$	$\omega = 10$
Variance	92%	90%	93%
Standard	90%	93%	100%
Exact	100%	100%	100%

Table 1. Percentages of p-values < 0.05 in 1000 simulated datasets by paired t-test of local variance and the standard and exact topological permutation tests. Computing time of one simulation with the exact test is on average 15000 folds faster than the standard test.

4. APPLICATION TO EPILEPTIC EEGS

Epilepsy is a neurological disorder marked by sudden recurrent episodes of sensory disturbance, loss of consciousness, or convulsions, associated with abnormal electrical activity in the brain. Refractory epilepsy affects one third of epilepsy patients worldwide, and the surgical treatment to help eliminate the condition depends on accurate localization of the seizure foci [6]. One of the important methods of treatment of refractory epilepsy is to remove the seizure foci after identifying that location in brain. Multiple approaches are used in determining the seizure foci. Here we apply the proposed exact permutation test of PLs to localize seizure foci.

4.1. Data.

EEG signals are sampled at the rate of 100 Hz from 8 channels (C3, C4, Cz, P3, P4, T3, T4, T5) in the central, temporal and parietal regions of the brain of a female patient [7, 8, 2]. Due to a lesion located on the cortical surface of the patient below the T3 channel, epileptic seizures are more likely to initiate from the left temporal lobe and abnormal electrical fluctuations are expected in EEG signals from channels around T3. During the entire EEG recording of 32,680 time points, a seizure initiates approximately halfway at the left temporal site (T3 channel). The raw and denoised EEG signals are shown in Figure 4. In subsequent analysis, the first and second phases of 16,340 time points in the EEG recording are designated as 'before' and 'during' seizure.

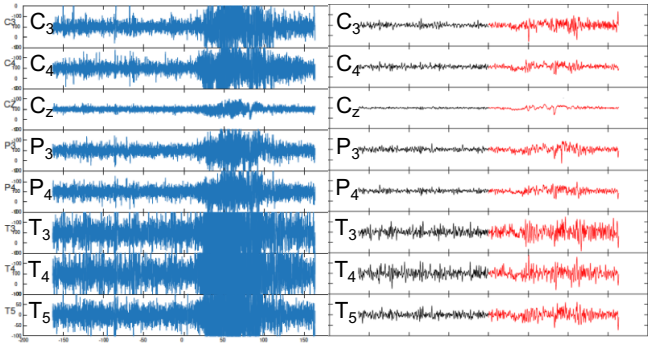

Fig. 4. Raw and denoised EEG signals.

4.2. Results.

Figure 5 shows the areas under the PLs A_ℓ^i for $i = 1, 2$ before and during seizure for denoised EEG signals. Consistent with results of the standard permutation test, the p-values of topological indifference reflect the spatial location of the seizure foci (T3, T5) as well as sites that may indicate seizure propagation (T4).

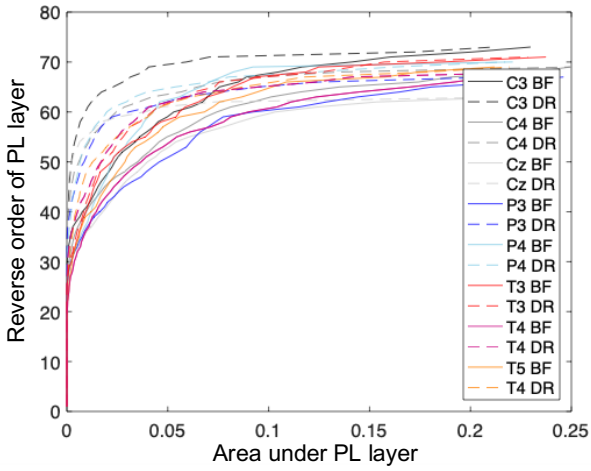

Fig. 5. The plot of areas under PLs before (BF) and during (DR) seizure for the 8 channels of EEG signals.

	Standard Test	Exact Test	Variance
C3	0.0001	0.0087	10^{-11}
C4	0.0014	0.1218	10^{-11}
Cz	0.0001	0.0019	10^{-11}
P3	0.0020	0.0282	10^{-11}
P4	0.0001	0.0338	10^{-11}
T3	0.0960	0.9999	10^{-11}
T4	0.0228	0.4662	10^{-8}
T5	0.1522	0.4755	10^{-11}

Table 2. Summary of p-values computed with all three tests. The p-values of the topological permutation tests consistently above the Bonferroni threshold of 0.05/8=0.0063 are shaded in gray.

5. DISCUSSION

We have shown the exact topological permutation test has much higher computing performance compared to the standard test. The current framework can be easily generalized to multi-trial EEG signals across patients to test the hypothesis that topological indfference marks the seizure foci and possible pathway of seizure propagation. Although the proposed method is motivated by seizure localization with EEG signals, the framework is widely applicable to other types of signals and application context such as semi-online detection of changes in local field potential (LFP) and functional magnetic resonance imaging (fMRI) signals in studies on induced stroke in rats, induced seizure, and cognitive studies on learning.

Acknowledgment. Support for Moo K. Chung was provided by the NIH Brain Initiative grant EB022856.

6. REFERENCES

[1] G. Carlsson, "Topology and data," *Bulletin of the American Mathematical Society*, vol. 46, pp. 255 – 308, 2009.

[2] Y. Wang, H. Ombao, and M.K. Chung, "Topological data analysis of single-trial electroencephalographic signals," *Annals of Applied Statistics*, vol. 12, pp. 1506–1534, 2018.

[3] P. Bubenik, "Statistical topological data analysis using persistence landscapes," *Journal of Machine Learning Research*, vol. 16, pp. 77 – 102, 2015.

[4] M.K. Chung, Z. Luo, A.L. Alexander, R.J. Davidson, and H.H. Goldsmith, "Exact combinatorial inference for brain images," *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, vol. 11070, pp. 629 – 637, 2018.

[5] S.L. Simpson, R.G. Lyday, S. Hayasaka, A.P. Marsh, and P.J. Laurienti, "A permutation testing framework to compare groups of brain networks," *Frontiers in Computational Neuroscience*, vol. 7, pp. 1–13, 2013.

[6] J.A. French, "Refractory epilepsy: clinical overview," *Epilepsia*, vol. 48, pp. 3 – 7, 2007.

[7] H.C. Ombao, J.A. Raz, R. von Sachs, and B.A. Malow, "Automatic statistical analysis of bivariate nonstationary time series," *Journal of the American Statistical Association*, vol. 96, pp. 543 – 560, 2001.

[8] H. Ombao, R. von Sachs, and W. Guo, "Slex analysis of multivariate nonstationary time series," *Journal of the American Statistical Association*, vol. 100, pp. 519 – 531, 2005.