Dirichlet depths for point process

Kai Qi[1] **and Yang Chen**[2] **and Wei Wu**[3]

[1]*Department of Statistics,*
Florida State University,
Tallahassee, FL 32306-4330
e-mail: kq15b@my.fsu.edu

[2]*Department of Statistics,*
Florida State University,
Tallahassee, FL 32306-4330
e-mail: cy16e@my.fsu.edu

[3]*Department of Statistics,*
Florida State University,
Tallahassee, FL 32306-4330
e-mail: wwu@stat.fsu.edu

Abstract: Statistical depths have been well studied for multivariate and functional data over the past few decades, but remain under-explored for point processes. A first attempt on the notion of point process depth was conducted recently where the depth was defined as a weighted product of two terms: (1) the probability of the number of events in each process and (2) the depth of the event times conditioned on the number of events by using a Mahalanobis depth. We point out that multivariate depths such as the Mahalanobis depth cannot be directly used because they often neglect the important ordering property in the point process events. To deal with this problem, we propose a model-based approach for point process systematically. In particular, we develop a Dirichlet-distribution-based framework on the conditional depth term, where the new methods are referred to as Dirichlet depths. We examine mathematical properties of the new depths and conduct asymptotic analysis. In addition, we illustrate the new methods using various simulated and real experiment data. It is found that the proposed framework provides a reasonable center-outward rank and the new methods have accurate decoding in one neural spike train dataset.

Keywords and phrases: Point process, Dirichlet depth, Poisson process, time warping, neural spike trains.

1. Introduction

Point process models have been well studied for many decades and widely applied in various disciplines, such as geography, seismology, astronomy, neuroscience, and so on. Those models are mainly focused on representing observations at each given time/location and have limited capability to measure the center-outward ranks of data. The center-outward rank, often referred to as statistical depth (depth for short), is a powerful tool to understand the features of underlying distribution such as spread and shape (Liu, Parelius and Singh, 1999). The study on depth has been focused on multivariate data and functional

data (Zuo and Serfling, 2000a; Lopez-Pintado and Romo, 2009; Mosler and Polyakova, 2012). In practice, depth has been successfully applied to address various practical problems such as classification (Lange, Mosler and Mozharovskyi, 2014), outlier detection (Chen et al., 2009), and diagnostics of nonnormality (Liu, Parelius and Singh, 1999).

The notion of statistical depth was first introduced and systematically studied on multivariate data by Tukey (1975). Since then, various definitions of multivariate depth have been proposed such as the convex hull peeling depth (Barnett, 1976), Oja depth (Oja, 1983), simplicial depth (Liu, 1990), Mahalanobis depth (Liu and Singh, 1993), and likelihood depth (Fraiman et al., 1999). As an axiomatic approach, a more general notion of depth for multivariate data was proposed by Zuo and Serfling (2000a), in which they summarized four desirable properties for multivariate depths, namely affine invariance, maximality at the center, monotonicity relative to the deepest point, and vanishing at infinity. In addition to multivariate data, depth for functional observations has received extensive attention in recent years (Lopez-Pintado and Romo, 2009; Mosler and Polyakova, 2012). Similar to the axiomatic approach in (Zuo and Serfling, 2000a), Nieto-Reyes and Battey (2016) provided a general definition of functional depth through six desirable properties, namely distance invariance, maximality at the center, decreasing with respect to the deepest point, upper semi-continuity in the function space, receptivity to convex hull with across the domain, and continuity in the probability measure.

Mathematical theories have also been extensively studied in majority of depth methods. For example, Nolan (1992) and Massé (2004) studied the convergence behavior of the halfspace depth and depth trimmed regions, and Koshevoy and Mosler (1997) studied the convergence behavior of the Zonoid depth. Furthermore, Dyckerhoff (2016) discussed the connections between different types of convergence for multivariate depths. Zuo and Serfling (2000b) studied the structural properties of trimmed regions, such as affine equivariance, nestedness, connectedness, and compactness.
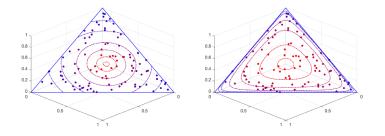
Our goal in this paper is to study the notion of statistical depth in temporal point process data. This is an under-explored area. The only previous work is given in (Liu and Wu, 2017), where the authors introduced the notion of depth in point process using a generalized Mahalanobis depth. Note that given the number of the events in a point process, the distribution of these events follow a multivariate framework. However, we point out that the multivariate depths cannot be directly used for point process data. This is because i) the number of events is a random variable, which is not described by the multivariate depths; ii) the events in a point process are an ordered sequence in a given (often finite) time domain. To the best of our knowledge, none of the multivariate depths studied the center-outward rank on ordered data.

Using mathematical notation, let $\mathbb{S}$ denotes the set of all point processes in a time domain $[T_1, T_2]$. Then an observed realization $s = (s_1, s_2, \cdots, s_k) \in \mathbb{S}$ can be treated as a vector in $\mathbb{R}^k$, where $|s| = k$ is the cardinality of s. This cardinality k can be any nonnegative integer. By the nature of temporal point process, the events $(s_1, s_2, \cdots, s_k)$ are ordered in a natural way as $T_1 \leq s_1 <$

$s_2 < \cdots < s_k \leq T_2$. Traditional depths defined on multivariate data neglect the importance of this order and will not be suitable for point process events. For example, suppose we inter-change the position of s_2 with s_1 and let $s' = (s_2, s_1, s_3, \cdots, s_k)$, traditional depth functions may still assign some positive depth values to s'. However, s' appears to be an outlier with zero probability, which is expected to have a zero depth value.

A depth function needs to take into account two types of randomness in a temporal point process S:(1) the number of events, or cardinality, in the process, denoted as $|S|$, and (2) the conditional distribution of these events given $|S|$. The notion of depth for point process was first studied by Liu and Wu (2017), where the authors defined a new depth framework as a weighted product of two terms: (1) the normalized probability of the number of events in each process and (2) the depth of the event times conditioned on the number of events by using the Mahalanobis depth. This weighted product is an appropriate way to address the two types of randomness. However, the Gaussian-kernel-based Mahalanobis conditional depth neglects the bounded and ordered property of the events. Here we use one example to illustrate how Gaussian-kernel-based conditional depth is inappropriate for the point process. The detailed simulation procedure is given in the Methods Section.

Basically, for a homogeneous Poisson process that only has two ordered events s_1 and s_2 in the time interval $[0, 1]$, the three inter-event times are: s_1, $s_2 - s_1$ and $1 - s_2$. These three intervals are nonnegative with the sum being 1, and therefore form a 2-dimensional simplex (i.e. a triangle) as shown in Figure 1.

(a) Mahalanobis depth contours   (b) Proposed depth contours

Fig 1: Example of the Mahalanobis depth and propose depth for HPP conditioned on 2 events in $[0, 1]$. The contours from outside to the center are with depth values 0.2, 0.3, 0.5, 0.7, 0.9, and 0.99, respectively. (a) Depth contours on inter-event times using the Mahalanobis Depth. (b) Same as (a) except using the proposed depth.

For the Mahalanobis depth, a Gaussian kernel is applied on the events, and therefore the inter-event times will also be represented by a Gaussian model. Typical Mahalanobis depth contours on the inter-event times are elliptical, as shown in Figure 1(a). We can see that such contours are not appropriate for the center-outward tendency since (1) the elliptical contours do not match the

triangular domain, and (2) the points on the border of simplex will be still assigned positive depth values by the Mahalanobis depth. A more reasonable contour plot is shown in Figure 1(b), where the proposed depth contours are triangle-like and resemble the triangular shape of the domain. In this paper, we will focus on explicit forms of point process depths with such contours.

Defining center-outward ranks for point process observations is a timely and important research topic. The goal of this paper is to develop a new depth framework for point processes systematically. Based on Liu and Wu (2017), our proposed framework of depth function for point processes is also defined as a weighted product of two terms aforementioned. In this paper, we focus on introducing new conditional depth functions based on the Dirichlet distribution. We will then discuss the desired mathematical properties and asymptotic behaviors.

The rest of this paper is organized as follows: In Sec. 2, we elaborate on the definitions of the new depths and provide computational procedures to effectively estimate them. The properties of the proposed depths are discussed in this section as well, followed by a thorough analysis with simulations. We then study the asymptotics of the new depths in Sec. 3. In Sec 4, we apply the new depths to classification problems in a real neural spike train dataset. Finally, we discuss and summarize the work in Sec. 5. All mathematical details are given in the Supplementary Materials.

2. Methods

In this section, we will at first review basic notation and then propose our new conditional depths for temporal point process. Since the new conditional depth functions are based on the Dirichlet distribution, we refer to them as the Dirichlet depths.

2.1. Notation and depth definition

Let $\mathbb{S}$ denote the set of all point processes in the time domain $[T_1, T_2]$. For any non-negative integer k, let $\mathbb{S}_k = \{s \in \mathbb{S} \mid |s| = k\} = \{(s_1, \cdots, s_k) \in \mathbb{R}^k | T_1 \leq s_1 \leq \cdots \leq s_k \leq T_2\}$ denote the set of all point processes in $\mathbb{S}$ with cardinality $|s| = k$. Hence, $\mathbb{S} = \bigcup_{k=0}^{\infty} \mathbb{S}_k$. For any $s \in \mathbb{S}$, a depth function for point process is a map $D : \mathbb{S} \to \mathbb{R}^+$ (set of nonnegative real numbers), $s \to D(s)$. Although our study in this paper focuses on the *simple point process* in which any two events cannot occur simultaneously, we allow the "=" sign in the domain to better derive important mathematical properties. This *boundary* set of $\mathbb{S}_k$ is denoted as $B_k = \{(s_1, \cdots, s_k) \in \mathbb{S}_k | \text{ at least one equality holds in } T_1 \leq s_1 \leq \cdots \leq s_k \leq T_2\}$.

As we have emphasized in Introduction, there are two types of randomness in a point process: (1) the number of events in each process, and (2) the conditional distribution of these event times. In (Liu and Wu, 2017), the number of events is modeled by a normalized Poisson mass function and the event times are modeled

by a multivariate Gaussian distribution. The depth framework of a point process s is then defined as a weighted product of two terms – the normalized probability of having $|s|$ events and the conditional depth using the Mahalanobis depth.

To better characterize the center-outward rank in a point process, we modify this framework by defining the depth as a weighted product of the following two terms: (1) a normalized one dimensional depth of the number of events in each process, and (2) a multivariate depth on the ordered event times conditioned on the number of events. The formal definition is given as follows:

Definition 2.1. *Given a random point process $S \in \mathbb{S}$ on $[T_1, T_2]$ with respect to a probability measure P, denote $P_{|S|}$ as a probability measure on the cardinality $|S|$ and $P_{S||S|}$ as the probability measure on the ordered events S (conditioned on $|S|$). For a realization $s \in S$, if $P_{|S|}(|S| = |s|) > 0$, then we define its depth $D(s; P)$ as:*

$$D(s; P) = w(|s|; P_{|S|})^r D_c(s; P_{S||S|}) \tag{2.1}$$

where $w(|s|; P_{|S|}) = \frac{D_1(|s|; P_{|S|})}{\max_k D_1(|s|=k; P_{|S|})}$ is the normalized one dimensional depth base on the cardinality $|s|$. $r > 0$ is the weight parameter and $D_c(s; P_{S||S|})$ is the depth of s conditioned on $|s|$. If $P_{|S|}(|S| = |s|) = 0$, then we directly define $D(s; P) = 0$.

The first term $w(|s|; P_{|S|})$ only depends on the distribution of $|S|$, with r as a tuning (weight) hyperparameter to balance its importance relative to the second term $D_c(s; P_{S||S|})$. As r gets larger, $w(|s|; P_{|S|})$ becomes a more dominant factor in the depth value $D(s; P)$. Various parametric or non-parametric methods can be adopted to estimate $w(|s|; P_{|S|})$. In this paper, we adopt the one dimensional half-space depth $D_1(\cdot)$ in the first term $w(|s|; P_{|S|})$. That is,

$$D_1(|s|; P_{|S|}) = \min\{P_{|S|}(|S| \leq |s|), P_{|S|}(|S| \geq |s|)\}. \tag{2.2}$$

In practical use, $D_1(|s|; P_{|S|})$ and $w(|s|; P_{|S|})$ can be easily estimated from samples.

Our study in this paper focuses on the second term $D_c(s; P_{S||S|})$, which describes the conditional depth when the number of events $|s|$ is given. In principle, any multivariate depth can be used as the conditional depth for point process if we treat $s \in \mathbb{S}_{|s|}$ as an $|s|$ dimensional vector. **However, we point out that such an approach neglects two important conditions of point process on $[T_1, T_2]$: (1) the event times are constrained on $[T_1, T_2]$, and (2) there exists a natural order in the event time sequence.** To address this issue, rather than defining the conditional depth function on the original point process space, we propose to define the conditional depth using inter-event times.

2.2. Equivalent representation and desirable properties

The point processes we discussed are bounded and ordered, i.e. $T_1 \leq s_1 \leq s_2 \leq \cdots \leq s_k \leq T_2$. Applying multivariate depth functions directly on $\mathbb{S}_k$ as conditional depths will tend to neglect the boundedness and orderedness

conditions. We propose to use inter-event times to represent a point process such that these important conditions are naturally satisfied.

2.2.1. Representation using inter-event times

It is well known that the point process can be equivalently represented by the inter-event times (IET). Here the IET of a point process $s_1, s_2, \cdots, s_k$ on $[T_1, T_2]$ are given as $u_1 = s_1 - T_1, u_2 = s_2 - s_1, \cdots, u_k = s_k - s_{k-1}, u_{k+1} = T_2 - s_k$. The IET sequence $u = (u_1, u_2, \cdots, u_{k+1})$ has k degrees of freedom and in fact forms a k-dimensional simplex (scaled standard simplex) as:

$$X_k = \{(u_1, u_2, \cdots, u_{k+1}) \in \mathbb{R}^{k+1} : u_1 + u_2 + \cdots + u_{k+1} = T_2 - T_1,$$
$$u_i \geq 0, i = 1, 2, \cdots, k+1\}.$$

X_k is bounded by the IET *boundary* set $Y_k = \{(u_1, u_2, \cdots, u_{k+1}) \in X_k : u_i = 0$ for at least one $i \in \{1, 2, \cdots, k+1\}\}$. Consistent to the point process boundary set B_k, the points in Y_k indicate a realization which has either two events happening simultaneously or one event happening at time T_1 or T_2. Both situations indicate extreme realizations (often with zero probability density) of a point process.

Based on this IET representation, we look for a conditional depth defined on the X_k simplex. Notice that the normalized IET sequence $(\frac{u_1}{T_2-T_1}, \frac{u_2}{T_2-T_1}, \cdots, \frac{u_{k+1}}{T_2-T_1})$ has the constant sum of 1. Therefore, one apparent option for the depth is the density function of Dirichlet distribution, which is commonly used as a prior in Bayesian statistics. Here we review the Dirichlet distribution which will be used to derive our conditional depth function: The Dirichlet probability density function of order $m \geq 2$ with concentration parameter vector $\mathbf{a} = (a_1, a_2, \cdots, a_m) \in \mathbb{R}^m$ with $a_i > 0$, $i = 1, \cdots, m$, is given as:

$$f(x_1, x_2, \cdots, x_m; a_1, a_2, \cdots, a_m) = \frac{\Gamma(\sum_{i=1}^m a_i)}{\prod_{i=1}^m \Gamma(a_i)} \prod_{i=1}^m x_i^{a_i-1}. \qquad (2.3)$$

where $(x_1, x_2, \cdots, x_m)$ is in the standard $m - 1$ simplex, i.e. $\sum_{i=1}^m x_i = 1$ and $x_i \geq 0, i = 1, 2, \cdots, m$. This density function is denoted as $Dirichlet(\mathbf{a}, m)$.

2.2.2. Desirable properties of the conditional depth for point process

In statistical depth literature, Zuo and Serfling (2000a) and Nieto-Reyes and Battey (2016) proposed important and desirable properties for depth on multivariate and functional data, respectively. They further claim that a depth function should be defined through desirable properties. Motivated by this claim, we list and discuss four desirable properties for a conditional depth function for point process as follow.

- P-1, Continuity and vanishing at the boundary: A conditional depth for point process is a map from the simplex X_k to $\mathbb{R}^+$. Since event times

are continuous on the time domain, a minimal requirement for a proper conditional depth should be continuity. In addition, an ideal conditional depth for point process should vanish at the boundary of the simplex domain.

- P-2, Maximality at the central point: This may be the most logical one among all properties since the "center" (the central point given measures of centrality) must have a maximal depth in a center-outward rank. In general, the notion of "center" is either a point of symmetry or the population mean (a.k.a. center of mass).
- P-3, Monotonicity relative to the deepest point: This property is also intuitive as depth value should decrease from the central point in a center-outward trend.
- P-4, Scale and shift invariance: The scale and shift invariance is a special case of the affine invariance in multivariate depth. Basically, a good depth is expected to be invariant with respect to scaling and translation on the time domain.

Let $\mathcal{P}_k$ denote the collection of all conditional probability measures on $\mathbb{S}_k$ and $P_{S||S|=k}$ denote the conditional probability measure of a given random point process $S \in \mathbb{S}_k$. We formally define the conditional statistical depth function as:

Definition 2.2. *Let the mapping $D_c(\cdot; \cdot) : \mathbb{S}_k \times \mathcal{P}_k \to \mathbb{R}^+$ be bounded, nonnegative, and satisfy P-1 to P-4; that is, assume:*

(i) $D_c(s; P_{S||S|=k})$ is a continuous map from $\mathbb{S}_k$ to $\mathbb{R}^+$ and $D_c(s; P_{S||S|=k}) = 0$ for any $s \in B_k$;

(ii) $D_c(\theta_k; P_{S||S|=k}) = \sup_{s \in \mathbb{S}_k} D_c(s; P_{S||S|=k})$ holds for any $P_{S||S|=k} \in \mathcal{P}_k$ having central point θ_k given a measure of centrality;

(iii) For any $P_{S||S|=k} \in \mathcal{P}_k$ having deepest point θ_k, $D_c(s; P_{S||S|=k}) \leq D_c(\theta_k + \alpha(s - \theta_k); P_{S||S|=k})$ for any $s \in \mathbb{S}_k$, $\alpha \in [0, 1]$; and

(iv) For any scaling coefficient $a \in \mathbb{R}^+$ and translation $b \in \mathbb{R}$, $D_c(s; P_{S||S|=k}) = D_c(as + b; P_{aS+b||S|=k})$.

Then $D_c(\cdot; P_{S||S|=k})$ is called a conditional statistical depth for point process in $\mathbb{S}_k$.

In addition to the above four properties, we note that the variation of a point process must satisfy two conditions: 1) the events are in the domain $[T_1, T_2]$, and 2) the events remain the temporal order. Such variation can be properly described by the set of time warping functions, defined as a boundary-preserving diffeomorphism $\Gamma = \{\gamma : [T_1, T_2] \to [T_1, T_2] \mid \gamma(T_1) = T_1, \gamma(T_2) = T_2, \dot{\gamma} > 0\}$, where the dot indicates the first order derivative (Srivastava and Klassen, 2016). Similar to the affine invariance in the multivariate depth, the following time warping invariance is also a desirable property for a conditional depth.

P-5*, Time warping invariance: For a point process S with cumulative intensity function $\Lambda_S(\cdot)$, $D_c(s; P_{S||S|=k}, \Lambda_S) = D_c(\gamma(s); P_{S||S|=k}, \Lambda_S^\gamma)$ holds for all

time warping $\gamma \in \Gamma$, where $\Lambda_S^\gamma = \Lambda_S \circ \gamma^{-1}$ is the cumulative intensity after the time warping transformation.

The time warping essentially allows any order-preserving nonlinear transformation of events in the given time domain. One might view P-5* as rather too strict (in contrast to the linear variance in P-4). Hence, the property P-5* is not listed in Definition 2.2. In the following sections, we will discuss all above properties in proposed conditional depths.

2.3. Dirichlet depth for homogeneous Poisson process

We at first develop a Dirichlet depth for the most classical temporal point process – homogeneous Poisson process (HPP).

2.3.1. Definition

For an HPP with constant rate λ on $[T_1, T_2]$, the first term $w(|s|; P_{|S|})$ in Equation (2.1) is simply the normalized depth on the number of events, which follows a Poisson distribution with mean $\lambda(T_2 - T_1)$. The challenge therefore stays on the conditional depth $D_c(s; P_{S||S|})$. As we have discussed, defining conditional depth for HPP on its IET representation will address the natural order issue, and ideally, the conditional depth should satisfy P-1 to P-4.

Before we step into the formal definition of the Dirichlet depth, we first look at the connection between HPP and Dirichlet distribution. For an HPP, we have defined IET (u_i , $i = 1, \cdots, k + 1$) as mentioned earlier. Conditioned on the number of events k, the normalized IET ($\frac{u_i}{T_2 - T_1}$, $i = 1, 2, \cdots, k+1$) will satisfy two conditions: (1) They share the same support, a k-dimensional standard simplex, as the Dirichlet distribution (also true for any point process). (2) They follow a flat $Dirichlet(\{1, \cdots, 1\}, k+1)$ distribution, which is in fact a uniform distribution over the standard k-dimensional simplex. The detail proof is given in Part A of the Supplementary Materials.

With a slight modification on Equation (2.3), we formally propose the Dirichlet depth for an HPP when the number of events is given as follows:

Definition 2.3. *Let $s = (s_1, s_2, \cdots, s_k)$ in $[T_1, T_2]$ be an observed homogeneous Poisson process. Denote $s_0 = T_1$, $s_{k+1} = T_2$. The Dirichlet depth of s (given $|s| = k$) is defined as:*

$$D_c(s; P_{S||S|=k}) = (k + 1) \prod_{i=1}^{k+1} \left(\frac{s_i - s_{i-1}}{T_2 - T_1}\right)^{\frac{1}{k+1}} \tag{2.4}$$

In particular, we have $D_c(s; P_{S||S|=0}) = 1$.

Remark 1 (Alternative motivation): The Dirichlet depth has been motivated by a more *natural* depth contour in Introduction. Alternatively, we here provide an information-theoretic motivation for Equation (2.4): To simplify notation, we let $[T_1, T_2] = [0, 1]$. For an HPP, conditioned on k events, we know

$\mathbb{E}(u_i) = 1/(k+1), i = 1, \cdots, k+1$. Let $q = (q_1, \cdots, q_{k+1})$ be the IET representation of any given HPP with k event. Treating u and q as two discrete distributions, the Kullback-Leibler (K-L) Divergence of q from u is

$$d_{KL}(u||q) = \sum_{i=1}^{k+1} u_i \log \frac{u_i}{q_i}.$$

Taking the expectation based on the distribution of u, we have

$$\mathbb{E}(d_{KL}(u||q)) = \mathbb{E}(\sum_{i=1}^{k+1} u_i \log u_i - \sum_{i=1}^{k+1} u_i \log q_i) = c - \frac{1}{k+1} \sum_{i=1}^{k+1} \log q_i,$$

where c is a constant. If we define a depth function of q as

$$H(q) \propto \exp(-\mathbb{E}(d_{KL}(u||q))) = \exp(-c) \cdot \prod_{i=1}^{k+1} q_i^{\frac{1}{k+1}},$$

then this is in the same form as in Eqn. (2.4). By letting the maximum being 1, we have

$$H(q) = (k+1) \prod_{i=1}^{k+1} q_i^{\frac{1}{k+1}}.$$

Therefore, the defined Dirichlet depth in Eqn. (2.4) can actually be built from this information-theoretic derivation.

Remark 2: In Equation (2.4), we have set the concentration parameters of the Dirichlet distribution a_i as $1 + \frac{1}{k+1}$ for $i = 1, 2, \cdots, k+1$. This constant value makes the Dirichlet depth a concave function with maximum at the conditional mean. The scale constant $(k+1)$ ensures $D_c(s; P_{S||S|=k})$ has an onto map to $[0, 1]$. This normalization makes conditional depths comparable for observations across different number of events.

Remark 3: The definition of conditional depth is not unique – any increasing function on D_c in Definition 2.3 can lead to another form of depth. For example, we can take a power $\alpha > 0$ on the Dirichlet depth in Equation (2.4) in the following form:

$$D_c(s; P_{S||S|=k}, \alpha) = [(k+1) \prod_{i=1}^{k+1} (\frac{s_i - s_{i-1}}{T_2 - T_1})^{\frac{1}{k+1}}]^\alpha.$$

The value of α decides the concentration level of the depth – a large α leads to a concentrated depth values around the deepest point, whereas a small α leads to a more evenly distributed depth values in the domain. For example, we can let α be proportional to the inverse of the total variance of the data.

2.3.2. Properties

By the IET representation, the conditional Dirichlet depth is defined on a simplex X_k. For this compact domain, traditional centers in the notion of commonly used symmetries (such as A-symmetry, C-symmetry, or H-symmetry in (Zuo and Serfling, 2000a)) cannot be directly applied. For example, when $k = 2$, the standard simplex X_2 is shown in Fig. 2 and the IET sequence is uniformly distributed in this simplex. We can see that the most natural "center" should be $(1/3, 1/3, 1/3)$. However, we point out that none of the above mentioned symmetries can be directly applied. Even for the most relaxed H-symmetry, there are only three half-space separation lines that can have probability $1/2$ at each side.
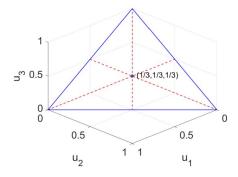

Fig 2: The 2-dimensional simplex with the IET representation on cardinality $|S| = 2$. The red lines denote the three median lines in the triangle where their intersection point is the center of mass $(1/3, 1/3, 1/3)$.

Hence, conditioned on $|S| = k$, we take the most "central point" as the common mathematical expectation. That is, the "central point" is:

$$\theta_k = \mathbb{E}(S \mid |S| = k) = (\mathbb{E}(S_1 \mid |S| = k), \cdots, \mathbb{E}(S_k \mid |S| = k)).$$

Using the derivation in Part A of the Supplementary Materials, the conditional expectation for HPP has the following closed-form:

$$\theta_k = \left(\frac{T_2 - T_1}{k + 1} + T_1, \frac{2(T_2 - T_1)}{k + 1} + T_1, \cdots, \frac{k(T_2 - T_1)}{k + 1} + T_1 \right).$$

The corresponding IET vector of conditional expectation is $(\frac{T_2 - T_1}{k+1}, \frac{T_2 - T_1}{k+1}, \cdots, \frac{T_2 - T_1}{k+1})$. On a k-dimensional simplex, this point is the same as the geometric center. For general point process other than the HPP, we will adopt the similar notion of "central point".

Theorem 2.1. *If the conditional mean is adopted as measurement of centrality, then $D_c(\cdot; P_{S||S|=k})$ in Definition 2.3 is a conditional statistical depth for homogeneous Poisson process in sense of Definition 2.2.*

The proof of Theorem 2.1 is given in Part B of the Supplementary Materials.

2.3.3. *Illustration*

In this subsection, we will at first examine the ranking performance of the Dirichlet depth on $\mathbb{S}_2$. We will then utilize the Dirichlet depth as the conditional depth in Equation (2.1) to study the ranking performance on 100 HPP realizations.

Conditioned on the cardinality $|S| = 2$, the inter-event times are uniformly distributed on a 2-dimensional simplex. Here we simulate 100 realizations from HPP conditional on $|S| = 2$ in time interval $[0, 1]$, and then apply both Dirichlet Depth for HPP (Equation (2.4)) and Mahalanobis depth for comparison. The result is shown in Figure 1 in the Introduction Section. We can see that compared with the (truncated) elliptic contours by the Mahalanobis depth, Dirichlet depth has smooth, triangle-like contours that are more compatible with the triangular IET domain. Moreover, the Mahalanobis depth assigns positive depth values to the points on the boundary. In contrast, the depth values on the boundary are all zero in the Dirichlet depth.

Next, we will apply Dirichlet depth for HPP (Equation (2.4)) as the conditional depth of the proposed depth framework (Equation (2.1)) on 100 HPP realizations in the interval $[0, 10]$. The detailed procedure is: (1) Randomly generate 100 HPP realizations in $[0, 10]$ with intensity rate $\lambda = 0.4$. (2) Compute the depth of cardinality $|s|$ of sampled realizations and normalize it as the first term $w(|s|; P_{|S|})$ in Equation (2.1). (3) Apply the Dirichlet depth for HPP as the conditional depth of Equation (2.1) for each realization. The result is shown in Figure 3.
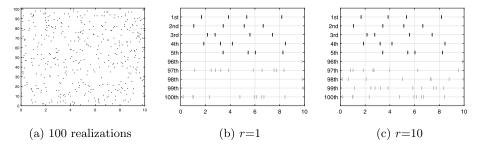

(a) 100 realizations   (b) r=1   (c) r=10

Fig 3: Ranking HPP realizations. (a) 100 HPP realizations on $[0, 10]$ with rate 0.4, where each row is one realization. (b) The top 5 and bottom 5 ranked realizations using the Dirichlet depth and $r = 1$. (c) Same as (b) except that $r = 10$.

We can see that the depth values depend on both $w(|s|; P_{|S|})$ and the conditional depth $D_c(s; P_{S||S|})$, and the ranks vary with different choice of r. The deepest realization is the one with the typical number of events ($|s| = 4$) and evenly distributed events (close to the center). When the value of r changes from 1 to 10, the first term $w(|s|; P_{|S|})$ becomes more dominant, so that realizations with cardinality close to 4 are more likely to be ranked on the top. Note that r is a hyperparameter in the depth definition. In practice, one can adjust r to

balance the w term and D_c term for different purposes. For example, as we illustrate in Appendix I, a cross-validation procedure may be applied to find an optimal value.

2.4. Dirichlet depth for general point process

Based on Definition 2.3 for the HPP, we can now define the Dirichlet depth for general point process.

2.4.1. Definition

Defining Dirichlet depth for general point process is more challenging since there is no direct connection between Dirichlet distribution and point process (other than the HPP). In this paper, we propose two different approaches for the definition: (1) naturally extend Equation (2.4) to general point process, and (2) transform the process to an HPP and then adopt Equation (2.4). At first, we extend the Dirichlet depth in Equation (2.4) to a general point process in $[T_1, T_2]$ by defining the center of IET as the conditional mean of the process. The formal definition is given as follows.

Definition 2.4. *Given the cardinality k, assume the conditional mean vector of a point process in time $[T_1, T_2]$ is $\mu_k = (\mu_{1,k}, \mu_{2,k}, \cdots, \mu_{k,k})$. For an observed realization $s = (s_1, s_2, \cdots, s_k)$, set $s_0 = \mu_{0,k} = T_1$ and $s_{k+1} = \mu_{k+1,k} = T_2$. If $\mu_{i,k} > \mu_{i-1,k}, i = 1, \cdots, k+1$, then the Dirichlet depth of s is defined as:*

$$D_c(s; P_{S||S|=k}) = \prod_{i=1}^{k+1} \left(\frac{s_i - s_{i-1}}{\mu_{i,k} - \mu_{i-1,k}} \right)^{\frac{\mu_{i,k} - \mu_{i-1,k}}{T_2 - T_1}}. \tag{2.5}$$

For an HPP S in $[T_1, T_2]$ conditioned on cardinality $|S| = k$, we have showed that its conditional mean is: $(\frac{T_2 - T_1}{k+1} + T_1, \frac{2(T_2 - T_1)}{k+1} + T_1, \cdots, \frac{k(T_2 - T_1)}{k+1} + T_1)$. In this case, it is easy to verify that Equation (2.5) is simplified to Equation (2.4). Therefore, Dirichlet depth for HPP is a special case of Equation (2.5). Note that the conditional depth value in Equation (2.5) only depends on the conditional expectation of the process with the same cardinality, which can be estimated by the conditional sample mean given a collection of realizations. Hence, the sample version of Dirichlet depth can be obtained by replacing its conditional means with sample means. That is, we can write the **sample Dirichlet depth** as:

$$D_c(s; P_{S||S|=k}^{(n_k)}) = \prod_{i=1}^{k+1} \left(\frac{s_i - s_{i-1}}{\overline{s}_{i,k}^{(n_k)} - \overline{s}_{i-1,k}^{(n_k)}} \right)^{\frac{\overline{s}_{i,k}^{(n_k)} - \overline{s}_{i-1,k}^{(n_k)}}{T_2 - T_1}}, \tag{2.6}$$

where $\overline{s}_k^{(n_k)} = (\overline{s}_{1,k}^{(n_k)}, \overline{s}_{2,k}^{(n_k)}, \cdots, \overline{s}_{k,k}^{(n_k)})$ are the estimated sample means conditioned on the cardinality k of n_k observed realizations and $P_{S||S|=k}^{(n_k)}$ denotes the empirical conditional probability measure.

Remark 4: Similar information-theoretic motivation in Remark 1 can be built for the definition of depth in Equation (2.5): Letting $[T_1, T_2] = [0, 1]$, the K-L

Divergence of any IET sequence q from u is still $d_{KL}(u||q) = \sum_{i=1}^{k+1} u_i \log \frac{u_i}{q_i}$. As $\mathbb{E}(u_i) = \mu_{i,k} - \mu_{i-1,k}$,

$$\mathbb{E}(d_{KL}(u||q)) = \mathbb{E}(\sum_{i=1}^{k+1} u_i \log u_i - \sum_{i=1}^{k+1} u_i \log q_i) = c - \sum_{i=1}^{k+1} (\mu_{i,k} - \mu_{i-1,k}) \log q_i,$$

where c is a constant. If we define a depth function of q as

$$H(q) \propto \exp(-\mathbb{E}(d_{KL}(u||q))) = \exp(-c) \cdot \prod_{i=1}^{k+1} q_i^{\mu_{i,k} - \mu_{i-1,k}}$$

This is in the same form as in Equation (2.5). By letting the maximum being 1, we have

$$H(q) = \prod_{i=1}^{k+1} (\frac{q_i}{\mu_{i,k} - \mu_{i-1,k}})^{\mu_{i,k} - \mu_{i-1,k}}.$$

Remark 5: The same α-power transformation on the Dirichlet depth for HPP (shown in Remark 3) can be applied to the Dirichlet depth in Equation (2.5) and sample Dirichlet depth in Equation (2.6).

2.4.2. Properties

For point processes in general, we also adopt the "central point" as the conditional mean given cardinality $|S| = k$. That is, $\theta_k = \mathbb{E}(S \mid |S| = k)$.

Theorem 2.2. *If the conditional mean is adopted as measurement of centrality, then $D_c(s; P_{S||S|=k})$ in Definition 2.5 is a conditional statistical depth for general point process in sense of Definition 2.2.*

The detailed proof of Theorem 2.2 is omitted due to its similarity to the HPP case. Now we examine the property P-5* of the Dirichlet depth. Let $S = (S_1, \cdots, S_k)$ be a random point process realization with k events. Then the conditional mean is $\mathbb{E}(S \mid |S| = k) = (\mathbb{E}(S_1 \mid |S| = k), \cdots, \mathbb{E}(S_k \mid |S| = k))$. Under the time warping $\gamma \in \Gamma$, a point process $s = (s_1, \cdots, s_k)$ will become $\gamma(s) = (\gamma(s_1), \cdots, \gamma(s_k))$. Similarly, the conditional mean will become $\mathbb{E}(\gamma(S) \mid |S| = k) = (\mathbb{E}(\gamma(S_1) \mid |S| = k), \cdots, \mathbb{E}(\gamma(S_k) \mid |S| = k))$. To simplify the notation on the conditional means, we let $\mu_{i,k} = \mathbb{E}(S_i \mid |S| = k)$ and $\mu_{i,k,\gamma} = \mathbb{E}(\gamma(S_i) \mid |S| = k), i = 1, \cdots, k$ and $\mu_{0,k} = \mu_{0,k,\gamma} = T_1, \mu_{k+1,k} = \mu_{k+1,k,\gamma} = T_2$. If we include the conditional means in the definition, the Dirichlet depth on the transformed point process is:

$$D_c(\gamma(s); P_{S||S|=k}) = \prod_{i=1}^{k+1} \left(\frac{\gamma(s_i) - \gamma(s_{i-1})}{\mu_{i,k,\gamma} - \mu_{i-1,k,\gamma}} \right)^{\frac{\mu_{i,k,\gamma} - \mu_{i-1,k,\gamma}}{\gamma(T_2) - \gamma(T_1)}}$$

$$\neq \prod_{i=1}^{k+1} (\frac{s_i - s_{i-1}}{\mu_{i,k} - \mu_{i-1,k}})^{\frac{\mu_{i,k} - \mu_{i-1,k}}{T_2 - T_1}} = D_c(s; P_{S||S|=k}).$$

The inequality holds because the time warping in general is a nonlinear transformation. Therefore, P-5* does not hold for the Dirichlet depth in Definition 2.4.

2.4.3. Bootstrapping estimation

Dirichlet depth in Equation (2.5) relies heavily on the conditional means. For point process in general, there are no closed-forms for the conditional means if conditional intensity function is unknown. In practice, given a set of point process realizations, we can apply the sample version Equation (2.6) to estimate the Dirichlet depth. However, for a given training data set, the sample size usually is not sufficiently large to properly estimate the conditional mean for each cardinality. Here we propose a bootstrapping approach to address this issue.

Given a data set of point process realizations $p_1, p_2, \cdots, p_n$, where p_i is a vector in $\mathbb{R}^{|p_i|}$ for $i = 1, 2, \cdots, n$. In general, those vectors do not have the same dimension, and therefore it is not possible to take an average to compute the conditional sample mean as we need in sample Dirichlet depth. We propose a bootstrap method to resample each realization p_i such that the resampled realizations p_i' has the desired dimension k. Then we can effectively estimate the conditional sample mean given cardinality $|S| = k$ by simply taking an average. The detailed steps are listed in Algorithm 1.

Basically, for each point process, we either remove events from it or add events via resampling from the overall data (excluding the given process) such that it has k events at the end. Then the conditional means can be easily estimated via all processes with k events. This procedure is under a basic assumption that the temporal events are history-independent (such as the commonly used Poisson process). The effectiveness of this algorithm is illustrated via simulation examples in Part C of the Supplementary Materials. Although this algorithm is based on a significant simplification, it performs reasonably well in practice. We will show the use of Algorithm 1 in a real experimental data in Section 4.

2.5. Alternative Dirichlet depth for general point process

The time warping transformation allows all events in a point process move within the given domain, while remaining the order of them. Ideally, the center-outward ranks of a set of point processes will be invariant if the same transformation is applied on all of them. However, we have shown that the Dirichlet depth in Equation (2.5) does not hold such invariance. In this subsection, we seek for an alternative definition to satisfy this property.

2.5.1. Definition

Note that we have defined the depth for HPP. For any point process, if we can find a way to transform it to an HPP, then the notion of Dirichlet depth can

---

Algorithm 1 Bootstrapping method to estimate conditional means

---

Require: Given a sequence of realizations of point process $p_1, p_2, \cdots, p_n$

Combine all events of $p_1, p_2, \cdots, p_n$ together, $p_{com} = \{p_1, p_2, \cdots, p_n\}$

for $k = 1$ *to* $max(|p_i|)$ **do**

 for $i = 1$ *to* n **do**

  if $|p_i| \geq k$, then uniformly randomly delete $|p_i| - k$ events in p_i

  Otherwise, add $k - |p_i|$ by uniformly re-sampling from other realizations (in $p_{com} - p_i$)

 end for

 Denote n resampled realizations as $p'_{1,k}, p'_{2,k}, \cdots, p'_{n,k}$, and then the estimated conditional mean is:

$$\overline{s}_k^{(n)} = \frac{1}{n} \sum_{i=1}^{n} p'_{i,k}$$

end for

return $[\overline{s}_k^{(n)}]_{k=1}^{max(|p_i|)}$

---

be directly applied. Actually, such transformation can be done using the well-known **time re-scaling theorem** (Brown et al., 2001). Basically, assuming the conditional probability of more than one event in $(t, t + \Delta t]$ given the history dependence H_t is 0 (this is also referred to as orderliness assumption), the theorem states that any point process with an integrable conditional intensity function can be converted into an HPP (Papangelou, 1972; Karr, 1991): Let $T_1 < s_1 < s_2 < \cdots < s_k \leq T_2$ be a realization from a point process with a conditional intensity function $\lambda(t|H_t) > 0$ for all $t \in (T_1, T_2]$. Then, the sequence $\Lambda_S(s_i) = \int_{T_1}^{s_i} \lambda(t|H_t)dt, i = 1, \cdots, k$ is a Poisson process with the unit rate in $(0, \Lambda(T_2)]$.

By applying this theorem, the notion of Dirichlet depth can be extended to general point processes. For a point process with known conditional intensity function, we can apply the time re-scaling theorem to convert it into an HPP in $[0, 1]$, and then use Equation (2.4) to compute its Dirichlet depth. Here we propose an alternative definition of the Dirichlet depth, referred to as time-rescaling-based (or TR-based) Dirichlet depth, as follows:

Definition 2.5. *For a point process S satisfying orderliness assumption in $[T_1, T_2]$ with conditional intensity function $\lambda(t \mid H_t) > 0$ and $\Lambda_S(t) = \int_{T_1}^{t} \lambda(u \mid H_u)du$, we define a time-rescaling-based Dirichlet depth of a realization $s = (s_1, s_2, ..., s_k)$ as:*

$$D_{c-TR}(s; P_{S||S|=k}, \Lambda_S) = (k+1) \prod_{i=1}^{k+1} \left(\frac{\Lambda_S(s_i) - \Lambda_S(s_{i-1})}{\Lambda_S(T_2)}\right)^{\frac{1}{k+1}}, \qquad (2.7)$$

where $s_0 = T_1$ and $s_{k+1} = T_2$,

Remark 6: The same α-power transformation on Dirichlet depth for HPP (shown in Remark 3) can be applied to the TR-based Dirichlet depth in Equation (2.7).

We can verify that the sequence $(\frac{\Lambda_S(S_1)}{\Lambda_S(T_2)}, \cdots, \frac{\Lambda_S(S_k)}{\Lambda_S(T_2)})$ follows an HPP in $[0, 1]$ with intensity $\Lambda_S(T_2)$. For point processes without history dependence such as

an inhomogeneous Poisson process (IPP), the time re-scaled HPP realization will be distributed in a deterministic time interval $(0, \Lambda_S(T_2)]$. However, for point processes with history dependence, the re-scaled HPP realizations have different time lengths. A normalization to the interval $[0, 1]$ by dividing $\Lambda_S(T_2)$ will help make comparison across realizations. Note that Definition 2.5 is not IET-based with respect to the original point process, and therefore there is no notion of simplex. This is fundamentally different from Definitions 2.3 and 2.4.

Moreover, for an HPP in $[T_1, T_2]$ with constant rate λ, $\Lambda_S(t) = \lambda(t - T_1)$. Then Equation (2.7) is simplified to

$$(k+1) \prod_{i=1}^{k+1} (\frac{\lambda(s_i - s_{i-1})}{\lambda(T_2 - T_1)})^{\frac{1}{k+1}} = (k+1) \prod_{i=1}^{k+1} (\frac{s_i - s_{i-1}}{T_2 - T_1})^{\frac{1}{k+1}}.$$

Therefore, the TR-based Dirichlet depth also generalizes Definition 2.3 for HPP.

Comparing Definitions 2.5 and 2.4, we can see that the depth in Definition 2.4 seems easier to use in practice as its depth value only depends on the conditional means. In contrast, in Definition 2.5 the randomness of a point process is due to the shape of its conditional intensity function, and TR-based Dirichlet depth entirely depends on this function. Hence, if the conditional intensity function is known, the proposed TR-based definition is expected to have an effective center-outward ranking. However, the conditional intensity function is often unknown in practical use. Indeed, perhaps the most challenging part of the TR-based Dirichlet depth is to properly estimate the conditional intensity.

Under different assumptions, various approaches for conditional intensity estimation have been proposed such as Peristimulus Time Histogram, Spline, and Inhomogeneous Markov Interval (Brown et al., 2001). The former two approaches assume that the underlying process is history-independent. The last one introduces a certain degree of history dependency (Markovian transition) in the model.

2.5.2. Properties

We will now examine the five properties for the TR-based Dirichlet depth $D_{c-TR}(s; P_{S||S|=k}, \Lambda_S)$ in Equation (2.7). We can easily verify that this depth satisfies P-1 (see proof in Part D of Supplementary Materials). In addition, this depth will be preserved under scale and shift transformation as defined in P-4 (the proof is trivial).

The function $\Lambda_S(\cdot)$ varies with respect to each point process. Hence, in general, there could be multiple maxima in the TR-based Dirichlet depth and a unique center will not exist (see one example in Part E of Supplementary Materials). Therefore, P-2 (and therefore P-3) will not be satisfied by the TR-based Dirichlet depth.

Under the time warping $\gamma \in \Gamma$, a point process $s = (s_1, s_2, \cdots, s_k)$ will become $\gamma(s) = (\gamma(s_1), \gamma(s_2), \cdots, \gamma(s_k))$. Similarly, we find that the transformed cumulative conditional intensity function $\Lambda_S^{\gamma} = \Lambda_S \circ \gamma^{-1}$. If we include the

cumulative conditional intensity function in the definition of Dirichlet depth, we have

$$D_{c-TR}(\gamma(s); P_{S||S|=k}, \Lambda_S^\gamma) = (k+1) \prod_{i=1}^{k+1} (\frac{\Lambda_S^\gamma(\gamma(s_i)) - \Lambda_S^\gamma(\gamma(s_{i-1}))}{\Lambda_S^\gamma(\gamma(T_2))})^{\frac{1}{k+1}},$$

$$= (k+1) \prod_{i=1}^{k+1} (\frac{\Lambda_S(s_i) - \Lambda_S(s_{i-1})}{\Lambda_S(T_2)})^{\frac{1}{k+1}} = D_{c-TR}(s; P_{S||S|=k}, \Lambda_S).$$

The detailed proof is given in Part F of the Supplementary Materials.

As a summary, we list the properties of all proposed Dirichlet depths in Table 1, where "T" denotes "true" and "F" denotes "false". For comparative purpose, we also include the properties of the Mahalanobis depth Liu and Wu (2017).

TABLE 1
Properties of the Proposed Dirichlet depths

Method	P-1	P-2	P-3	P-4	P-5*
Dirichlet Depth on HPP	T	T	T	T	F
Dirichlet Depth on Point Process	T	T	T	T	F
TR-based Dirichlet Depth on Point Process	T	F	F	T	T
Mahalanobis Depth on Point Process	F	T	T	T	F

2.5.3. Illustration

In this subsection, we at first examine the ranking performance of the Dirichlet depths on $\mathbb{S}_2$. We will then utilize the Dirichlet depth and TR-based Dirichlet depth as the conditional depth in Equation (2.1) to study the ranking performance on 100 inhomogeneous Poisson process (IPP) realizations.

Conditioned on the cardinality $|S| = 2$, we simulate 100 realizations from IPP with intensity function $\lambda(t) = t^3$ in time interval $[0, 1]$, and then apply the sample Dirichlet Depth (Equation (2.6)) and Mahalanobis depth for comparison. The result is shown in Figure 4. Again, we can see that the Dirichlet depth contours fit the domain more reasonably.

We now illustrate the ranking performance using the proposed Dirichlet depths. 100 random IPP realizations are generated on $[0, 2\pi]$ with intensity function $\lambda(t) = 1 - cos(t)$. This intensity and the 100 realizations are shown in Figure 5(a). The total intensity is $\Lambda_S = \int_0^{2\pi} \lambda(t)dt = 2\pi$, and therefore the depth of the cardinality $D_1(|s|; P_{|S|})$ reaches its maximum at $|S| = 6$.

Given a set of a point processes, we need to estimate the conditional means in order to apply the sample Dirichlet depth and need to estimate the intensity function for TR-based Dirichlet depth. In this example, we use Algorithm 1 to estimate the conditional means. Under the Poisson assumption, the intensity function can be easily estimated with training samples. In this example, we use Equation (2.2) to compute depth on the number of events $w(|s|; P_{|S|})$. We set weight coefficient r to two different values of 1 and 0.01. The ranking result is shown in Figure 5.

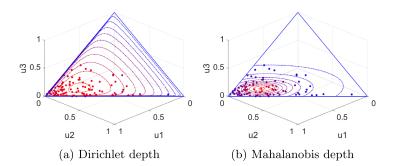(a) Dirichlet depth   (b) Mahalanobis depth

Fig 4: Example of Dirichlet depth and Mahalanobis depth for IPP conditioned on 2 events in $[0, 1]$. The contours from outside to the center are with depth values 0.08, 0.18, 0.28, 0.38, 0.48, 0.58, 0.68, 0.78, 0.88, and 0.98, respectively. (a) Depth contours and IETs using the Dirichlet Depth. (b) Depth contours and IETs using the Mahalanobis Depth.
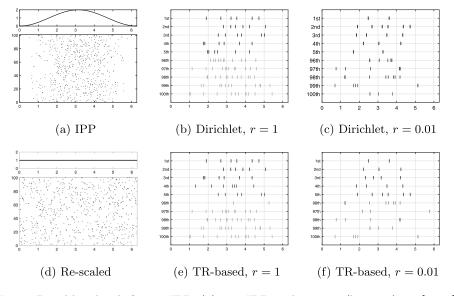

(a) IPP   (b) Dirichlet, $r = 1$ (c) Dirichlet, $r = 0.01$

(d) Re-scaled   (e) TR-based, $r = 1$ (f) TR-based, $r = 0.01$

Fig 5: Dirichlet depth for an IPP. (a) 100 IPP realizations (bottom) on $[0, 2\pi]$ with the intensity function $\lambda(t) = 1 - cos(t)$ (top). (b) The top 5 and bottom 5 ranked realizations with the sample Dirichlet depth and $r = 1$. (c) Same as (b) except for $r = 0.01$. (d) TR-based transferred realizations (bottom) and its intensity function (top). (e) Same as (b) except for the TR-based Dirichlet depth. (f) Same as (e) except for $r = 0.01$.

Comparing Panels (b) and (e) (where $r = 1$), we can see that the ranking results based on sample Dirichlet depth function and TR-based Dirichlet depth function are very similar – four out of five top-ranked realizations are the same.

Four out of five bottom-ranked realizations are the same as well. We also compare Panels (c) and (f) where $r = 0.01$. Although the overall ranks changed dramatically from where $r = 1$, both methods agree on four out of five deepest, and four out of five shallowest realizations.

3. Asymptotic theory

In this section, we will investigate the asymptotic behavior of the sample depth function for point process based on our proposed framework (Equation (2.1)), where the sample and population conditional depths are given in Equations (2.6) and (2.5), respectively. For $s \in \mathbb{S}_k$, we need to estimate the first term $w(|s| = k; P_{|S|})$ and Dirichlet depth $D_c(s; P_{S||S|=k})$. r is a pre-set hyperparameter. Given a sample set $\mathcal{S}^{(n)}$ that contains n realizations from a point process on interval $[T_1, T_2]$, $w(|s| = k; P_{|S|})$ can be estimated by empirical probability measure $P_{|S|}^{(n)}$:

$$D_1(|s| = k; P_{|S|}^{(n)}) = \min\left(\frac{\#\{|\mathcal{S}^{(n)}| \leq k\}}{n}, \frac{\#\{|\mathcal{S}^{(n)}| \geq k\}}{n}\right),$$

where $\#\{|\mathcal{S}^{(n)}| \leq (\geq)k\}$ denotes the number of processes in $\mathcal{S}^{(n)}$ with less (more) than or equal to k events. Hence,

$$w(|s| = k; P_{|S|}^{(n)}) = \frac{D_1(|s| = k; P_{|S|}^{(n)})}{\max_{0 \leq g \leq K} D_1(|s| = g; P_{|S|}^{(n)})}$$

with a pre-determined $K \in \mathbb{N}$. Basically, $w(|s| = k; P_{|S|}^{(n)})$ is the halfspace depth base on empirical probability mass function, normalized with maximum being 1. The conditional Dirichlet depth proposed in Equation (2.5) can be estimated by the sample Dirichlet depth in Equation (2.6). Then we have a sample version of Equation (2.1) $D(s; P^{(n)})$ in the following form:

$$D(s; P^{(n)}) = w(|s| = k; P_{|S|}^{(n)})^r D_c(s; P_{S||S|=k}^{(n)}) \tag{3.1}$$

To simplify the theoretical derivation, we make the following two assumptions.

1. The number of events in each process has a constant upper bound $K \in \mathbb{N}$, which can be arbitrarily large.
2. If $P_{|S|}(|S| = k) > 0$, then the conditional means satisfy $\mu_{i,k} > \mu_{i-1,k}, i = 1, \cdots, k+1, k = 0, 1, \cdots, K$.

We have defined $\mathbb{S}_k = \{s = (s_1, \cdots, s_k) \in \mathbb{R}^k | T_1 \leq s_1 \leq \cdots \leq s_k \leq T_2\}$. Let $E^{(K)} = \bigcup_{k=0}^{K} \mathbb{S}_k$. The depth function in Equation (2.1) is a function $D : E^{(K)} \rightarrow [0, 1]$. Our main asymptotic result is given as follows:

Theorem 3.1. *For arbitrarily large $K \in \mathbb{N}$, let $s \in E^{(K)}$ be a point process realization in the time domain $[T_1, T_2]$. If the two assumptions given above are satisfied, then*

$$\sup_{s \in E^{(K)}} |D(s; P^{(n)}) - D(s; P)| \to 0 \ a.s. \ (as \ n \to \infty) \tag{3.2}$$

Furthermore, for $\alpha \in (0,1)$, denote $D^\alpha \equiv \{s \in E^{(K)} \mid D(s; P) \geq \alpha\}$ and $D_n^\alpha \equiv \{s \in E^{(K)} \mid D(s; P^{(n)}) \geq \alpha\}$ as $\alpha - trimmed$ regions. Then for any $\epsilon \in (0, \min\{\alpha, 1 - \alpha\})$,

1. *$D_n^{\alpha+\epsilon} \subset D^\alpha \subset D_n^{\alpha-\epsilon}$ for n sufficiently large.*
2. *$D_n^\alpha \to D^\alpha$ a.s. as $n \to \infty$ if $P(\{s \in E^{(K)} \mid D(s; P) = \alpha\}) = 0$.*

The proof of Theorem 3.1 is given in Part G of the Supplementary Materials.

4. Real data: Geniculate ganglion spike trains

For application of the proposed depth framework, one may use the depth values to classify point processes. In this section, we will illustrate such use in decoding of neural spike train data. More in-depth simulation studies can be found in Part H in the Supplementary Materials.

Basically, we will use a spike train dataset to demonstrate the classification performance of the proposed framework, where spike trains can be naturally treated as a temporal point process. This dataset was previously used in Lawhern et al. (2011). In the experiment, adult male Sprague-Dawley rat's geniculate ganglion tongue neurons were stimulated by 6 different solutions: KCI (salty), CA (sour), NaCl (salty), QHCI (bitter), MSG (umami) and Sucr (sweet) for 10 times each. The experiment consists of three time periods: 2-second pre-stimulus period, 2.5-second stimulus application period and 2-second post-stimulus period.

For illustrative purposes, we only use spike trains in the stimulus application period and the post-stimulus period, and only select two typical neurons cells: one electrolyte generalist cell and one acid generalist cell. For each cell, we take 5 spike trains for each of 6 different tastes to train parameters in the depth functions (i.e. as the training set), and another 5 spike trains to perform classification task (i.e. as the test set). That is, 60 spike trains are been selected for each cell. The spike trains of the training set with respect to the 6 different solutions from those two cells are shown in Figure 6.

For comparative purpose, we also study the classification rates for the generalized Mahalanobis depth and likelihood method Liu and Wu (2017). To use the likelihood method and TR-based Dirichlet depth, one needs to estimate the conditional intensity functions. Here we take a common assumption that the underlying processes for spike train observations are Poisson processes, and therefore the intensity functions can be obtained via conventional kernel smoothing methods. For the Dirichlet depth and generalized Mahalanobis depth, their parameters can be estimated via the bootstrapping in Algorithm 1. Once the

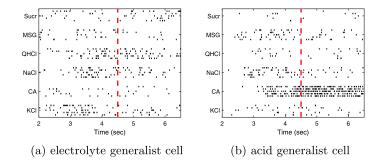(a) electrolyte generalist cell   (b) acid generalist cell

Fig 6: Sample spike trains for different cells. (a) Sample spike trains of an electrolyte generalist cell. Time interval before the vertical dashed line is the stimulus application period and that after is the post-stimulus period. (b) Same as (a) except for an acid generalist cell.

models are fit, we can then classify a test spike train to one of the six tastes in which the depth value or likelihood is the highest.

In addition, one can smooth the point process observations using kernel functions, and then adopt multivariate depth to conduct the center-outward ranking. In this dataset, we use a Gaussian kernel function and the classical modified bandwidth to estimate the depths. We then conduct classification with these depth values. The classification (or decoding) accuracy rates of all the five methods are summarized in Table 2.

TABLE 2
Comparison of decoding performance

Method	electrolyte generalist cell	acid generalist cell
Dirichlet depth	**0.73**	**0.83**
TR-based Dirichlet depth	**0.73**	**0.86**
Generalized Mahalanobis depth	0.70	0.76
Modified Bandwidth depth	0.50	0.63
Likelihood method	0.47	0.33

We can see that the proposed Dirichlet depth and TR-based Dirichlet depth have very good accuracies (between 0.73 and 0.86 for the two cells) in classifying test spike trains to one of the six types of tastes (a random guess only results in $\frac{1}{6} = .17$ accuracy). Such accuracies are better than those by the likelihood method, the modified bandwidth depth, and Mahalanobis depth (between 0.33 and 0.76). In particular, given the extremely small sample size (5 in each taste), the classification performance of the proposed methods is highly successful. We point out that the TR-based Dirichlet depth depends largely on the estimated intensity function and the sample size in this example is too small to have a robust estimate. Nevertheless, this TR-based Dirichlet method still has accurate classification with respect to the noisy intensity estimation. In contrast, the

other intensity-based approach, the likelihood method, results in much lower classification rate.

5. Summary

In this paper, we have proposed a new framework to measure depth for point process observations. The proposed depths include three components: 1) normalized depth on the number of events, 2) conditional depth given the number of events, and 3) the weight parameter. Our study focuses on the definitions of these new depths and examines their mathematical properties. The depth is at first defined for the classical homogeneous Poisson process by using the equivalent inter-event time representation. For general point process, we propose two different definitions: one is a direct generalization on the homogeneous Poisson case, and the other is based on the well-known time re-scaling theorem. We examine the mathematical properties for each of these depths and provide a theoretical investigation on the asymptotic of the sample Dirichlet depth. Moreover, we apply the proposed depth functions to a neural decoding problem. The result indicates that the proposed framework provides a proper center-outward rank and the new methods have accurate classification performance.
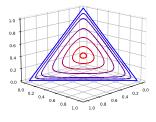

Fig 7: The Mahalanobis depth on the ILR transformed inter-event times for HPP conditioned on 2 events in $[0, 1]$. The contours from outside to the center are with depth values 0.2, 0.3, 0.5, 0.7, 0.9, and 0.99, respectively.

We have compared the proposed depths with the recently developed Mahalanobis depth on point process and demonstrated the superiority of the new methods. However, we point out that the Mahalanobis depth is directly used on the IET simplex, which naturally produces inappropriate center-outward ranks on the observations. Based on the classical theory on compositional data analysis, observations on a multi-dimensional simplex can be bijectively transformed to a Euclidean space via certain nonlinear transformations such as ALR (additive logratio transform), CLR (center logratio transform), and ILR (isometric logratio transform) (Aitchison, 1986; Egozcue et al., 2003). In particular, ILR is a preferred method because it is symmetric with respect to variables, and its covariance is of full rank. For example, Figure 7 shows the contours of the Mahalanobis depth after the ILR transformation on a homogeneous Poisson process, and the shape of contours are very similar to our proposed depth shown in Figure 1(b). These transformations provide a new paradigm of the notion of depth

for point processes, and we will study this topic systematically and thoroughly in our future work.

We point out that the Dirichlet depth is a new approach to define the conditional depth for point process. To the best of our knowledge, no other methods have been proposed to study this problem. More in-depth topics, such as the shape of depth contours and trimmed regions in a high dimension, can be further explored. For practical application, we have only investigated the classification performance by the proposed depth framework. Other applied topics, such as clustering and outliers detection, can also be studied in the future.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data,Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd. MR0865647

Barnett, V. (1976). The ordering of multivariate data. *Journal of the Royal Statistical Society. Series A (General)* **139** 318–355. MR0445726

Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E. and Frank, L. M. (2001). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation* **14** 325–346.

Chen, Y., Dang, X., Peng, H. and Bart, H. L. J. (2009). Outlier detection with the kernelized spatial depth function. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31** 288–305.

Dyckerhoff, R. (2016). Convergence of depths and depth-trimmed regions. *arXiv preprint arXiv:1611.08721*.

Egozcue, J. J., Pawlowsky-Glahn, V., Figueras, G. and Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology* **35** 279–300. MR1986165

Fraiman, R., Meloche, J., García-Escudero, L., Gordaliza, A., He, X., Maronna, R., Yohai, V., Sheather, S., Mckean, J., G. Small, C. and Wood, A. (1999). Multivariate L-estimation. *Test* **8** 255–317.

Karr, A. (1991). *Point Processes and Their Statistical Inference, Second Edition,. Probability: Pure and Applied*. Taylor & Francis. MR1113698

Koshevoy, G. and Mosler, K. (1997). Zonoid trimming for multivariate distributions. *The Annals of Statistics.* **25** 1998–2017. MR1474078

Lange, T., Mosler, K. and Mozharovskyi, P. (2014). Fast nonparametric classification based on data depth. *Statistical Papers* **55** 49–69. MR3152767

Lawhern, V., A. A, N., Wu, W. and Contrares, R. J. (2011). Spike rate and spike timing contributions to coding taste quality information in rat periphery. *Frontiers in Integrative Neuroscience* **5** 1–14.

Liu, R. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics* **18** 405–414. MR1041400

Liu, R. Y., Parelius, J. M. and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by Liu and Singh). *The Annals of Statistics* **27** 783–858. MR1724033

LIU, R. Y. and SINGH, K. (1993). A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association* **88** 252–260. MR1212489

LIU, S. and WU, W. (2017). Generalized Mahalanobis depth in point process and its application in neural coding. *The Annals of Applied Statistics* **11** 992–1010. MR3693555

LOPEZ-PINTADO, S. and ROMO, J. (2009). On the concept of depth for functional data. *Journal of American Statistical Association* **104** 718–734. MR2541590

MASSÉ, J.-C. (2004). Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli* **10** 397–419. MR2061438

MOSLER, K. and POLYAKOVA, Y. (2012). General notions of depth for functional data. *Manuscript* **1981**.

NIETO-REYES, A. and BATTEY, H. (2016). A topologically valid definition of depth for functional data. *Statistical Science* **31** 61–79. MR3458593

NOLAN, D. (1992). Asymptotics for multivariate trimming. *Stochastic Processes and their Applications* **42** 157–169. MR1172513

OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statistics & Probability Letters* **1** 327–332. MR0721446

PAPANGELOU, F. (1972). Integrability of expected increments of point processes and a related random change of scale. *Transactions of the American Mathematical Society – TRANS AMER MATH SOC* **165**. MR0314102

SRIVASTAVA, A. and KLASSEN, E. P. (2016). *Functional and Shape Data Analysis.* Springer-Verlag New York. MR3821566

TUKEY, J. W. (1975). Mathematics and picturing data. *International Congress of Mathematicians* **2** 523-531. MR0426989

ZUO, Y. and SERFLING, R. (2000a). General notions of statistical depth function. *The Annals of Statistics* **28** 461–482. MR1790005

ZUO, Y. and SERFLING, R. (2000b). Structural properties and convergence results for contours of sample statistical depth functions. *The Annals of Statistics* **28** 483–499. MR1790006

Appendix A: Relationship between HPP and Dirichlet distribution

Suppose $\{s_1, s_2, \cdots, s_k\}$ (unordered) is a sequence from a homogeneous Poisson process with constant rate λ in time interval $[0, 1]$ (without loss of generality). We know that given the number of events (cardinality) $|s| = k$, the unordered events $\{s_1, s_2, \cdots, s_k\}$ are independently and uniformly distributed on $[0, 1]$. Let $s = (s_{(1)}, s_{(2)}, \cdots, s_{(k)})$ denote the increasingly ordered set of $\{s_1, s_2, \cdots, s_k\}$. If we define the inter-event times as $u_1 = s_{(1)}, u_2 = s_{(2)} - s_{(1)}, \cdots, u_k = s_{(k)} - s_{(k-1)}, u_{k+1} = 1 - s_{(k)}$, then conditional on cardinality $|s| = k$, we know that the p.d.f. of the ordered statistics s is

$$f_s(s_{(1)}, s_{(2)}, \cdots, s_{(k)}) = k!, 0 \le s_{(1)} \le s_{(2)} \le \cdots \le s_{(k)} \le 1.$$

By the change of variables and the fact that $u_{k+1} = 1 - \sum_{i=1}^{k} u_i$, the p.d.f. of inter-event time vector $(u_1, u_2, \cdots, u_{k+1})$ is:

$$f_u(u_1, u_2, \cdots, u_{k+1}) = k!, \text{ where } \sum_{i=1}^{k+1} u_i = 1, u_i \geq 0, i = 1, \cdots, k+1.$$

Therefore, conditioned on cardinality $|s| = k$, the vector $(u_1, u_2, \cdots, u_{k+1})$ follows a Dirichlet distribution (2.3) with concentration parameter ($\alpha_i = 1, i = 1, 2, \cdots, k+1$).

Furthermore, it is well known that the $i-$th order statistics of k uniformly distributed event $s_{(i)}$ follows a Beta distribution in the form:

$$s_{(i)} \sim Beta(i, \ k+1-i)$$

Hence, for $i = 1, 2, \cdots, k+1$, the expectation of $s_{(i)}$ is $\mathbb{E}s_{(i)} = \frac{i}{k+1}$. The expectation of i-th inter-event time u_i is:

$$\mathbb{E}(u_i) = \mathbb{E}(s_{(i)} - s_{(i-1)}) = \mathbb{E}(s_{(i)}) - \mathbb{E}(s_{(i-1)}) = \frac{i}{k+1} - \frac{i-1}{k+1} = \frac{1}{k+1}.$$

Appendix B: Proof of properties of Dirichlet depth for HPP

Suppose $S = (S_1, S_2, \cdots, S_k)$ is a homogeneous Poisson process with rate λ in time interval $[T_1, T_2]$. Then conditional on cardinality $|S| = k$, the Dirichlet depth in Equation (2.4) for s satisfy:

(1) continuity and vanishing at the boundary,
(2) maximality at the conditional mean (the center),
(3) monotonicity relative to the deepest point,
(4) scale and shift invariant

Proof: (1) By the definition of the boundary for a point process conditional on cardinality $|s| = k$, we can formally write the vanishing at boundary property of Dirichlet depth as:

$$D_c(s; P_{S||S|=k}) = 0 \text{ iff } \exists i \in \{1, 2, \cdots, k+1\}, s.t. \ s_i = s_{i-1},$$

where as $s_0 = T_1$ and $s_{k+1} = T_2$. Based on the closed-form definition in Equation (2.4), it is easy to see the continuity and vanishing at the boundary for the defined Dirichlet depth.

(2) Conditioned on the cardinality $|s| = k$, finding the maximum for $D_c(s; P_{S||S|=k})$ in Equation (2.4) is equivalent to finding maximum for the function:

$$f(x) = \sum_{i=1}^{k+1} \log(x_i),$$

where $x = (x_1, \cdots, x_{k+1})$, $x_i = \frac{s_i - s_{i-1}}{T_2 - T_1}$, and $\sum_{i=1}^{k+1} x_i = 1$.
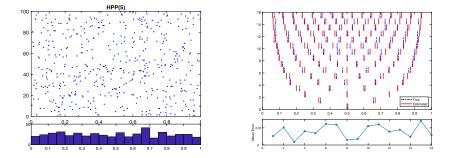
Using the Lagrange multiplier, it is easy to obtain the unique, global maximum $\{x_i = \frac{1}{k+1}, \; i = 1, 2, \cdots, k+1\}$, which equals the conditional mean of IET (the center).

(3) It is easy to find that the Hessian matrix of the function $f(x)$ is negative definite. This implies that $D_c(s; P_{S||S|=k})$ is a log-concave function on the domain $\mathbb{S}_k$. Combining this result with the global maximum in Property (2), we can see that $D_c(s; P_{S||S|=k})$ satisfies Property (3).

(4) The proof is trivial and omitted here.

Appendix C: Illustration of the proposed bootstrap method

In this section, we will illustrate the effectiveness of the proposed bootstrap method (Algorithm 1) through two simulation examples. In the first simulation, 100 realizations are simulated from an HPP with constant rate $\lambda = 5$ in $[0, 1]$. The simulated realizations and the histogram of events over all realizations are shown in Figure 8 (a). Given the cardinality $k = 1, 2, \cdots$, or 16, the true conditional means have a closed vector form $(\frac{1}{k+1}, \cdots, \frac{k}{k+1})$. We then apply Algorithm 1 to estimate the conditional means. A comparison of the true and estimated conditional means is shown in Figure 8 (b). We can see that the estimation is accurate for the number of events k varying from 1 to 16 (average errors are only around 0.01 in the time domain $[0, 1]$).
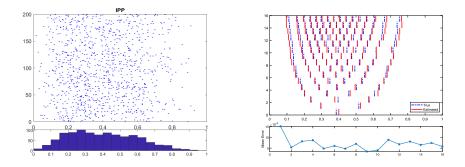

(a) Simulated HPP realizations with histogram

(b) True and estimated conditional means

Fig 8: Simulation One to illustrate Algorithm 1. (a) Simulated HPP realizations with $\lambda = 5$ in $[0, 1]$ (top panel) and the histogram of all events (bottom panel). (b) Plot of true (blue lines) and estimated (red lines) conditional means given k varying from 1 to 16 (top panel) and plot of mean error (average of the errors at all events) at different values of k (bottom panel).

Similarly, in the second simulation, we follow the same procedures as in the first example except that the realizations are simulated from an IPP with conditional distribution $f(S = (s_1, \cdots, s_k) \mid |S| = k) = \prod_{i=1}^{k} Beta(s_i; \alpha = 2, \beta = 3)$. In this case, there are no closed-form for the true means, but we can approximate

them with conventional Monte-Carlo method (using additional ample simulations from the true model). The result is shown in Figure 9, where the mean errors are also around 0.01 in the time domain $[0, 1]$.

(a) Simulated IPP realizations with histogram

(b) True and estimated conditional means

Fig 9: Simulation Two to illustrate Algorithm 1. (a) Simulated IPP realizations with conditional density $f(s \mid |S| = k) = \prod_{i=1}^{k} Beta(s_i; \alpha = 2, \beta = 3)$ (top panel) and its histogram (bottom panel). (b) Plot of true (blue lines) and estimated (red lines) conditional means given k varying from 1 to 16 (top panel) and plot of mean error at different values of k (bottom panel).

Appendix D: Proof of continuity of TR-based Dirichlet depth

Proof: For a point process $s = (s_1, s_2, \cdots, s_k) \in \mathbb{S}_k$ with an integrable and continuous intensity function $\lambda(t \mid H_t)$ such that: $0 < \lambda(t \mid H_t) \leq M$ for all $t \in [T_1, T_2]$ and for some finite M. Denote the re-scaled sequence after applying time re-scaling theorem as $\{\Lambda_S(s_i) : i = 1, 2, \cdots, k\}$. Then

$$\Lambda_S(s_1) = \int_{T_1}^{s_1} \lambda(t)dt$$

$$\Lambda_S(s_2) = \Lambda_S(s_1) + \int_{s_1}^{s_2} \lambda(t \mid H_{s_1})dt$$

$$\cdots$$

$$\Lambda_S(s_k) = \Lambda_S(s_{k-1}) + \int_{s_{k-1}}^{s_k} \lambda(t \mid H_{s_{k-1}})dt$$

In order to prove the TR-based Dirichlet depth is continuous, we only need to show that $\Lambda(\cdot)$ is a continuous map. Consider another sequence $s' = (s'_1, s'_2, \ldots s'_k) \in \mathbb{S}_k$ near s. Based on the assumption on $\lambda(t \mid H_t)$,

$$|\Lambda_S(s_1) - \Lambda_S(s'_1)| = |\int_{s'_1}^{s_1} \lambda(t)dt| \leq M|s_1 - s'_1| \to 0 \quad (s'_1 \to s_1)$$

Moreover,

$$|\Lambda_S(s_2) - \Lambda_S(s_2')| \leq |\int_{s_1}^{s_2} \lambda(t|H_{s_1})dt - \int_{s_1'}^{s_2'} \lambda(t|H_{s_1'})dt| + |\Lambda_S(s_1) - \Lambda_S(s_1')|$$

As $\lambda(t \mid H_t)$ is a continuous map on $s = (s_1, ...s_k)$, $\lim_{s_1' \to s_1} \lambda(t|H_{s_1'}) = \lambda(t|H_{s_1})$. Therefore,

$$\lim_{(s_1', s_2') \to (s_1, s_2)} |\Lambda_S(s_2) - \Lambda_S(s_2')|$$

$$\leq \lim_{(s_1', s_2') \to (s_1, s_2)} |\int_{s_1}^{s_2} \lambda(t|H_{s_1})dt - \int_{s_1'}^{s_2'} \lambda(t|H_{s_1'})dt| + |\Lambda_S(s_1) - \Lambda_S(s_1')|$$

$$\leq \lim_{(s_1', s_2') \to (s_1, s_2)} (M|s_2 - s_2'| + M|s_1 - s_1'| + M|s_1 - s_1'|) = 0$$

Similarly, we can prove the components function $\{\Lambda_S(s_1), \Lambda_S(s_2), \cdots, \Lambda_S(s_k)\}$ are all continuous functions on $(s_1, s_2, \cdots, s_k)$, and therefore, the TR-based Dirichlet Depth is also continuous on s.

Appendix E: Example of existence of multiple maximal points

It is easy to see that the TR-based Dirichlet depth reaches maximum when $\frac{\Lambda_S(s_i) - \Lambda_S(s_{i-1})}{\Lambda_S(T_2)} = \frac{1}{k+1}, i = 1, 2, \cdots, k+1$. However, we point out that for history-dependent point process, there could be multiple realizations $s \in \mathbb{S}_k$ such that $\frac{\Lambda_s(s_i) - \Lambda_s(s_{i-1})}{\Lambda_s(T_2)} = \frac{1}{k+1}, i = 1, 2, \cdots, k+1$. We will provide one example as follows.

We consider a point process s on $[0, 1]$ with cardinality $|s| = 1$. Denote the unique event as τ. Given the conditional intensity function $\lambda(u \mid H_u)$, the cumulative intensity of s can be denoted as:

$$\Lambda_s(t) = \int_0^t \lambda(u)du, \text{ for } t \leq \tau$$

$$\Lambda_s(t) = \int_0^\tau \lambda(u)du + \int_\tau^t \lambda(u \mid \tau)du, \text{ for } t > \tau$$

Let $s = (\tau^*)$ denote the maximizer of the TR-based depth with cardinality 1. That is,

$$\Lambda_s(\tau^*) = \frac{1}{2}\Lambda(1).$$

Then,

$$\int_0^{\tau^*} \lambda(u)du = \int_{\tau^*}^1 \lambda(u \mid \tau^*)du$$

If we further set its conditional intensity function as:

$$\lambda(t) = 1/c \cdot t^{1/c-1}, c > 0 \text{ for } t \leq \tau, \text{ and}$$

$$\lambda(t \mid \tau) = 16/9(1/2 + \tau) \text{ for } t > \tau.$$

Then

$$\int_0^{\tau^*} \lambda(u)du = (\tau^*)^{1/c}, c > 0$$

$$\int_{\tau^*}^1 \lambda(u \mid \tau^*)du = \frac{16}{9}(\frac{1}{2} + \tau^*)(1 - \tau^*)$$

The two functions $f_1(t) = t^{1/c}, f_2(t) = \frac{16}{9}(\frac{1}{2} + t)(1 - t)$ are shown in Fig. 10 below. Note that $f_1(t)$ is in the range $[0, 1]$ for $t \in [0, 1]$ and the quadratic $f_2(t)$ has maximum 1 when $t = 1/4$. We can see that when c is sufficiently large, there are multiple solutions for the equation $f_1(t) = f_2(t)$ (three solutions are shown in Fig. 10(b)). Therefore, we have proven that there could be multiple maximizers for the TR-based depth.
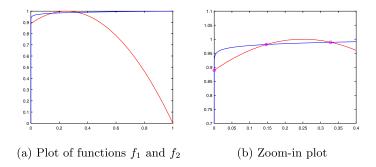

(a) Plot of functions f_1 and f_2 (b) Zoom-in plot

Fig 10: (a) Plot of functions $f_1(t) = t^{1/c}, c = 100$ (in blue) and $f_2(t) = \frac{16}{9}(\frac{1}{2} + t)(1 - t)$ (in red) on $[0, 1]$. (b) Zoom-in of (a) for the three intersection points.

Appendix F: Proof of time warping invariance of time-rescaling based Dirichlet depth

Consider a point process $S \in \mathbb{S}_k$ in $[0, T]$ (without loss of generality) with integrable function $\lambda(t \mid H_t)$. Define a time warping map of S as $\gamma \in \Gamma, \Gamma = \{\gamma : [0, T] \to [0, T] \mid \gamma(0) = 0, \gamma(T) = T, \dot{\gamma} > 0\}$. For a realization $s = (s_1, s_2, \cdots, s_k) \in S$ such that $0 \le s_1 < s_2 < \cdots < s_k \le T$, the TR-based Dirichlet depth on s is invariant under time warping, i.e. $D_{c-TR}(s; P_{S||S|=k}, \Lambda_S) = D_{c-TR}(\gamma(s); P_{S||S|=k}, \Lambda_S^\gamma)$.

Proof: Note that the time warping function γ is a strictly increasing, bijective, and differentiable function. After the time warping, the corresponding process becomes $\gamma(s) = (\gamma(s_1), \gamma(s_2), \cdots, \gamma(s_k))$ with $0 \le \gamma(s_1) < \gamma(s_2) < \cdots < \gamma(s_k) \le T$. To study the Dirichlet depth after the time warping, we at first

examine the conditional intensity function in the warped process. By definition, the transformed conditional intensity function is

$$\lambda_\gamma(t \mid H_t)$$
$$= \lim_{\Delta t \to 0} \frac{\mathbb{P}(\text{one event in } [t, t + \Delta t] \text{ in the transformed process} \mid H_t)}{\Delta t}$$
$$= \lim_{\Delta t \to 0} \frac{\mathbb{P}(\text{one event in } [\gamma^{-1}(t), \gamma^{-1}(t + \Delta t)] \text{ in the original process} \mid H_{\gamma^{-1}(t)})}{\Delta t}$$
$$= \lim_{\Delta t \to 0} \frac{\mathbb{P}(\mathbb{N}(\gamma^{-1}(t + \Delta t)) - \mathbb{N}(\gamma^{-1}(t)) = 1 \mid H_{\gamma^{-1}(t)})}{\Delta t}$$
$$= \lim_{\Delta t \to 0} \frac{\mathbb{P}(\mathbb{N}(\gamma^{-1}(t + \Delta t)) - \mathbb{N}(\gamma^{-1}(t)) = 1 \mid H_{\gamma^{-1}(t)})}{\gamma^{-1}(t + \Delta t) - \gamma^{-1}(t)} \cdot \frac{\gamma^{-1}(t + \Delta t) - \gamma^{-1}(t)}{\Delta t}$$
$$= \lambda(\gamma^{-1}(t) \mid H_{\gamma^{-1}(t)}) \dot\gamma^{-1}(t)$$

Then the cumulative conditional intensity function in the transformed process is

$$\Lambda_S^\gamma(t) = \int_0^t \lambda_\gamma(t \mid H_t) dt$$
$$= \int_0^t \lambda(\gamma^{-1}(t) \mid H_{\gamma^{-1}(t)}) \dot\gamma^{-1}(t) dt$$
$$= \int_0^{\gamma^{-1}(t)} \lambda(u \mid H_u) du \quad (u = \gamma^{-1}(t))$$
$$= \Lambda_S(\gamma^{-1}(t))$$

Hence, $\Lambda_S^\gamma = \Lambda_S \circ \gamma^{-1}$ and $\Lambda_S^\gamma \circ \gamma = \Lambda_S$. By the definition of TR-based Dirichlet depth as Equation (2.7), we have

$$D_{c-TR}(\gamma(s); P_{S||S|=k}, \Lambda_S^\gamma) = (k+1) \prod_{i=1}^{k+1} \left(\frac{\Lambda_S^\gamma(\gamma(s_i)) - \Lambda_S^\gamma(\gamma(s_{i-1}))}{\Lambda_S^\gamma(\gamma(T))} \right)^{\frac{1}{k+1}},$$

$$= (k+1) \prod_{i=1}^{k+1} \left(\frac{\Lambda_S(s_i) - \Lambda_S(s_{i-1})}{\Lambda_S(T)} \right)^{\frac{1}{k+1}} = D_{c-TR}(s; P_{S||S|=k}, \Lambda_S)$$

Appendix G: Proof of the asymptotic theorem

Proof: (Part I: proof of uniform convergence on the depth function):
First, we need to prove Equation (3.2) in Theorem 3.1. That is, the sample depth $D(s; P^{(n)})$ defined as Equation (3.1) uniformly converges to depth function in Equation (2.1) almost surely. Given a set of independent point processes $\{s_j\}_{j=1}^n$. By Assumption 1, $s_j \in E^{(K)}, j = 1, \cdots, n$. Then the first term $w(\cdot)$ in Equation (2.1) can be estimated as:

$$w(|s| = k; P_{|S|}^{(n)}) = \frac{D_1(|s| = k; P_{|S|}^{(n)})}{\max_{0 \leq g \leq K} D_1(|s| = g; P_{|S|}^{(n)})},$$

where

$$D_1(|s| = k; P_{|S|}^{(n)}) = \min(\frac{\#\{\{|s_j|\}_{j=1}^n \le k\}}{n}, \frac{\#\{\{|s_j|\}_{j=1}^n \ge k\}}{n}).$$

By the Strong Law of Large Numbers (SLLN), we can easily show that:

$$w(|s| = k; P_{|S|}^{(n)}) \to w(|s| = k; P_{|S|}) \text{ a.s. (as } n \to \infty)$$

If $P_{|S|}(|S| = |s|) = 0$, then apparently $P_{|S|}^{(n)}(|S| = |s|) = 0$. By Definition 2.1,

$$D(s; P^{(n)}) = D(s; P) = 0.$$

Then the uniform convergence naturally holds. Therefore, without loss of generality, we can further assume that for any $k \in \{0, \cdots, K\}, P_{|S|}(|S| = k) > 0$. We refer to this as **Assumption 3**.

Denote the processes $s_j = (s_{j,1}, \cdots, s_{j,|s_j|}), j = 1, \cdots, n$. Let n_k denote the number of processes in $\{s_j\}_{j=1}^n$ with k events and $\overline{s}_{i,k}^{(n_k)}$ denote their sample mean (corresponding to $\mu_{i,k}$), $k = 0, 1, \cdots, K$. By Assumption 3, $n_k \to \infty$ as $n \to \infty$. Then by the SLLN again, we have

$$\overline{s}_{i,k}^{(n_k)} \to \mu_{i,k} \text{ a.s. } i = 1, \cdots, k+1.$$

Therefore, for any $s \in E^{(K)}$, as $n \to \infty$,

$$D(s; P^{(n)}) \to D(s; P) \text{ (a.s.)}$$

Our goal is to show that this convergence is uniform on $E^{(K)}$. The proof has two steps: at first, we will show that for any given dimension $k \in \{0, 1, \cdots, K\}$,

$$\sup_{s \in S_k} |D(s; P^{(n)}) - D(s; P)| \to 0 \text{ a.s. as } n \to \infty$$

As $w(k; P_{|S|}^{(n)}) \to w(k; P_{|S|})$ (a.s.), we only need to show that $D_c(s; P_{S||S|=k}^{(n)})$ uniformly converges to $D_c(s; P_{S||S|=k})$ almost surely for $s \in \mathbb{S}_k$. By Assumption 2, there exists $\epsilon > 0$ such that $\mu_{i,k} - \mu_{i-1,k} > \epsilon$ for any $i \in \{1, \cdots, k+1\}, k \in \{1, \cdots, K\}$. We use $C > 0$ to denote a constant (independent of i, k and K). As the time domain is a finite interval $[T_1, T_2]$, we have

$$
\begin{aligned}
&\left| (\frac{s_i - s_{i-1}}{\overline{s}_{i,k}^{(n)} - \overline{s}_{i-1,k}^{(n)}})^{\frac{\overline{s}_{i,k}^{(n)} - \overline{s}_{i-1,k}^{(n)}}{T_2 - T_1}} - (\frac{s_i - s_{i-1}}{\mu_{i,k} - \mu_{i-1,k}})^{\frac{\mu_{i,d} - \mu_{i-1,d}}{T_2 - T_1}} \right| \\
\le &\left| (\frac{s_i - s_{i-1}}{\overline{s}_{i,k}^{(n)} - \overline{s}_{i-1,k}^{(n)}})^{\frac{\overline{s}_{i,k}^{(n)} - \overline{s}_{i-1,k}^{(n)}}{T_2 - T_1}} - (\frac{s_i - s_{i-1}}{\mu_{i,k} - \mu_{i-1,k}})^{\frac{\overline{s}_{i,k}^{(n)} - \overline{s}_{i-1,k}^{(n)}}{T_2 - T_1}} \right| \\
&+ \left| (\frac{s_i - s_{i-1}}{\mu_{i,k} - \mu_{i-1,k}})^{\frac{\overline{s}_{i,k}^{(n)} - \overline{s}_{i-1,k}^{(n)}}{T_2 - T_1}} - (\frac{s_i - s_{i-1}}{\mu_{i,k} - \mu_{i-1,k}})^{\frac{\mu_{i,d} - \mu_{i-1,d}}{T_2 - T_1}} \right|
\end{aligned}
$$

$$\leq C \left| (\frac{\mu_{i,k} - \mu_{i-1,k}}{\overline{s}_{i,k}^{(n)} - \overline{s}_{i-1,k}^{(n)}})^{\frac{\overline{s}_{i,k}^{(n)} - \overline{s}_{i-1,k}^{(n)}}{T_2 - T_1}} - 1 \right|$$

$$+ \left| (\frac{s_i - s_{i-1}}{\mu_{i,k} - \mu_{i-1,k}})^{\frac{\overline{s}_{i,k}^{(n)} - \overline{s}_{i-1,k}^{(n)}}{T_2 - T_1}} - (\frac{s_i - s_{i-1}}{\mu_{i,k} - \mu_{i-1,k}})^{\frac{\mu_{i,d} - \mu_{i-1,d}}{T_2 - T_1}} \right|$$

$$\rightrightarrows 0 \ (a.s.)$$

The uniform convergence holds in the first term as it is independent of $\{s_i\}$. In the second term, the base is the same in the two exponential expressions, and the difference is at the exponent. We can simply use a Lagrange's mean value theorem, and the uniform convergence can be obtained.

Note that the above uniform convergence is true for any $i \in \{1, \cdots, k+1\}$. As the Dirichlet depths $D_c(s; P_{S||S|=k}^{(n)})$ and $D_c(s; P_{S||S|=k}))$ are just a product of $k+1$ terms in the above form. We can easily obtain that

$$\sup_{s \in \mathbb{S}_k} |D_c(s; P_{S||S|=k}^{(n)}) - D_c(s; P_{S||S|=k}))| \to 0 \ a.s. \ as \ n \to \infty$$

In the second step, we will show this uniform convergence also holds on $E^{(K)}$. In fact, let Ω be the sample space and define the set:

$$A = \{\omega \in \Omega : \sup_{s \in E^{(K)}} |D(s, \omega; P^{(n)}) - D(s, \omega; P)| \to 0\}$$

$$= \{\omega \in \Omega : \sup_{s \in \bigcup_{k=0}^K \mathbb{S}_k} |D(s, \omega; P^{(n)}) - D(s, \omega; P)| \to 0\}$$

$$= \{\omega \in \Omega : \sup_{0 \leq k \leq K} \sup_{s \in \mathbb{S}_k} |D(s, \omega; P^{(n)}) - D(s, \omega; P)| \to 0\}$$

If we also let

$$A_k = \{\omega \in \Omega : \sup_{s \in \mathbb{S}_k} |D(s, \omega; P^{(n)}) - D(s, \omega; P)| \to 0\},$$

then $A^c \subset \bigcup_{k=1}^K A_k^c$. Using the result from Step 1, $P(A_k^c) = 0, k = 0, \cdots, K$. Hence:

$$P(A^c) \leq P(\bigcup_{k=0}^K A_k^c) \leq \sum_{k=0}^K P(A_k^c) = 0$$

Finally, we have proven that

$$\sup_{s \in E^{(K)}} |D(s; P^{(n)}) - D(s; P)| \to 0 \ a.s. \ as \ n \to \infty$$

(Part II: proof of convergence on the α-trimmed region): By the result in Part I, for any $\epsilon > 0$, there exists $N \in \mathbb{N}$, when $n > N$, $|D(s; P^{(n)}) - D(s; P)| < \epsilon$ almost surely for any $s \in E^{(K)}$. That is, with probability 1, for any $n > N$,

$$\{s \in E^{(K)} | D(s; P^{(n)}) \geq \alpha + \epsilon\} \subset \{s \in E^{(K)} | D(s; P) \geq \alpha\}$$

$$\subset \{s \in E^{(K)}|D(s; P^{(n)}) \geq \alpha - \epsilon\}$$

Therefore, $D_n^{\alpha+\epsilon} \subset D^\alpha \subset D_n^{\alpha-\epsilon}$ for n sufficiently large.

Taking upper limit on $D_n^{\alpha+\epsilon}$ and lower limit on $D_n^{\alpha-\epsilon}$, we have that with probability 1

$$\limsup_{n \to \infty} D_n^{\alpha+\epsilon} \subset D^\alpha \subset \liminf_{n \to \infty} D_n^{\alpha-\epsilon}.$$

Letting $\epsilon \to 0$, we have that with probability 1

$$\limsup_{n \to \infty}\{s \in E^{(K)}|D(s; P^{(n)}) > \alpha\} \subset D^\alpha \subset \liminf_{n \to \infty}\{s \in E^{(K)}|D(s; P^{(n)}) \geq \alpha\}.$$

If $P(\{s \in E^{(K)}|D(s; P^{(n)}) = \alpha\}) = 0$, then with probability 1,

$$\limsup_{n \to \infty} D_n^\alpha \subset D^\alpha \subset \liminf_{n \to \infty} D_n^\alpha.$$

That is, $D_n^\alpha \to D^\alpha$ a.s. as $n \to \infty$.

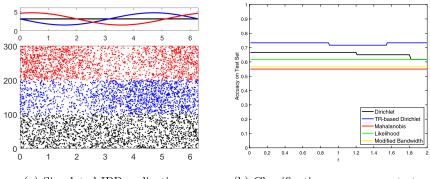## Appendix H: Simulations on classification study

In this appendix section, we will examine the classification performance of our proposed depth framework based on simulated Poisson processes, the most commonly used point process models for neuronal spike trains Brown et al. (2001). In this case, the intensity function is deterministic and can be effectively estimated from given data. For comparison, we will also study classification using the likelihood method and Mahalanobis depth values Liu and Wu (2017). We explore two examples in this study. In each example, we simulate multiple groups of Poisson processes, and then we classify them by the likelihood or depth values. The detailed procedures is given as follows:

1. Independently simulate 100 realizations for each group by its intensity functions, and then partition data into training and test processes (training:test = 80:20 in Example 1. training:test=70:30 in Example 2).
2. Estimate parameters in all methods (Dirichlet depth, TR-based Dirichlet depth, modified bandwidth depth, likelihood method, and Mahalanobis depth) for each group using the training set. In this process, we adopt Algorithm 1 to estimate the conditional means and covariances and apply nonparametric smoothing methods to estimate the intensity functions;
3. Based on the trained model, estimate the depth (or likelihood) for each test process, and assign the process to the group with largest depth (or likelihood) value.

Example 1: Differing in the event time distribution

In the first example, there are three groups of simulations on $[0, 2\pi]$. Their intensity functions are $\lambda_1(t) = \frac{20}{2\pi}$, $\lambda_2(t) = \frac{20(2-\sin(t))}{\int_0^{2\pi} 2-\sin(t)dt}$, and $\lambda_3(t) = \frac{20(2-\sin(t-2))}{\int_0^{2\pi} 2-\sin(t-2)dt}$,

respectively. These intensity functions are normalized, so that the total intensity $\Lambda(2\pi) = 20$ for all three groups. See Figure 11(a) for these three intensity functions and 100 simulations based on them, respectively. Under this simulation framework, the number of events in any process follows the same distribution $Poisson(20)$, and the first term $w(|s|)$ should be the same across groups. However, as the models are estimated from the training data, $w(|s|)$ may still slightly vary across groups. The classification result, mainly based on the event distributions, is shown in Figure 11(b). We see that the rate only slightly varies in the range of $[0, 2]$ for the coefficient r in each of three depth methods. As there is no such weight in the likelihood method and modified bandwidth depth method, their rates remain invariant in the range.

(a) Simulated IPP realizations   (b) Classification accuracy on test set

Fig 11: Classification on simulations in Example 1: (a) Three intensity functions $\lambda_1(t)$ (black), $\lambda_2(t)$ (blue), and $\lambda_3(t)$ (red) in the upper panel and 100 simulated processes with these functions in the low panel. (b) The classification rate of the likelihood method (green), Mahalanobis depth (red), Dirichlet depth (black), modified bandwidth depth (yellow) and TR-based Dirichlet depth (blue), under different weight coefficient r.

This example shows that the classification accuracy of the TR-based Dirichlet depth and Dirichlet depth are slightly better than that of the Mahalanobis depth, modified bandwidth depth, and likelihood. One basic problem for the likelihood method is that it cannot properly distinguish realization from an HPP as the likelihood will be a constant with respect to the event times. As the Mahalanobis method depends on covariance estimation, whereas 80 training processes may not provide sufficient samples for that purpose. This leads to the relatively worst performance of this method in this example.

Example 2: Differing in the number of event distribution

In the second example, we will let the distributions of the number of events vary across groups, whereas the distribution of the events will be the same.

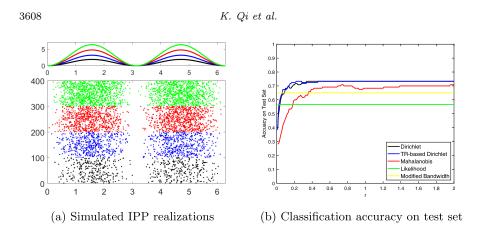(a) Simulated IPP realizations

(b) Classification accuracy on test set

Fig 12: Classification on simulations in Example 2: (a) Four intensity functions $\lambda_1(t)$ (black), $\lambda_2(t)$ (blue), $\lambda_3(t)$ (red), and $\lambda_4(t)$ (green) in the upper panel and 100 simulated processes with these functions in the low panel. (b) The classification rate of the likelihood method (green), Mahalanobis depth (red), Dirichlet depth (black), modified bandwidth depth (yellow) and TR-based Dirichlet depth (blue), under different weight coefficient r.

In particular, let the baseline "shape" density being $f(t) = \frac{1-\cos(2t)}{\int_0^{2\pi} 1-\cos(2t)dt}, t \in [0, 2\pi]$. The four intensity functions are: $\lambda_1(t) = 6f(t)$, $\lambda_2(t) = 10f(t)$, $\lambda_3(t) = 15f(t)$, and $\lambda_4(t) = 20f(t)$, respectively. See Figure 12(a) for these four intensity functions and 100 simulations based on them, respectively. Under this condition, the first term $w(|s|; P_{|S|})$ is more dominant as compared to the second term $D_c(s; P_{S||S|})$. The classification result is shown in Figure 12(b). We see that the rate varies in the range of $[0, 2]$ for the coefficient r in each of three depth methods. The rates of the likelihood method and the modified bandwidth depth are still constant in the range.

In this example, the two Dirichlet depths and Mahalanobis depth have similar classification performance. When r is very small, the distribution of the events will be the main factor. As this distribution is similar across different groups, we see an unsatisfactory classification rate at around 0.3-0.4 only. When the value of r gets larger, the accuracy improves. The peak is at around 0.75 for the two Dirichlet methods and around 0.70 for the Mahalanobis method. Theses accuracies remain stable for large values of r.

Appendix I: Illustration of choosing weight coefficient through cross-validation

This section is designed to illustrate how to choose the weight coefficient r (in Equation 2.1) for classification tasks through a K-Fold cross-validation. The data generation method is the same as what we discussed in section H. Here, we simulate four groups of Poisson processes on $[0, 2\pi]$ with intensity func-

tions: $\lambda_1(t) = \frac{6(1-\cos(2t))}{\int_0^{2\pi} 1-\cos(2t)dt}$, $\lambda_2(t) = \frac{20(2-\sin(t))}{\int_0^{2\pi} 2-\sin(t)dt}$, $\lambda_3(t) = \frac{20(2-\sin(t-2))}{\int_0^{2\pi} 2-\sin(t-2)dt}$, and $\lambda_4(t) = \frac{20(1-\cos(2t))}{\int_0^{2\pi} 1-\cos(2t)dt}$.

The detailed procedure can be summarized as follows:

1. Independently simulate 100 realizations for each group, and then partition data into K equal-sized folds (each fold contains $4 \times 100/K$ realizations).
2. Choose one fold as the validation set and the remaining $K-1$ folds as the training set. On the training set, we adopt Algorithm 1 to estimate the conditional means for Dirichlet depth and apply nonparametric smoothing methods to approximate the intensity functions for the TR-based Dirichlet depth.
3. For realizations in the validation set, estimate the first probability term $w(\cdot)$, Dirichlet depth, and TR-based Dirichlet depth based on the trained models (of 4 groups). Then, set $r = 0$, compute the depth as defined in Equation 2.1, and assign the process to the group with the largest depth value.
4. Sequentially increase r by 0.01 and record the classification accuracy of the validation set for each r. Stop the process when $r = 2$.
5. Repeat Steps $2 - 4$ for K times, with each of the K folds used as the validation set.
6. The optimal value of weight coefficient r is the one that maximizes the average classification accuracy over K runs.

(a) Simulated IPP realizations   (b) Classification accuracy on validation set
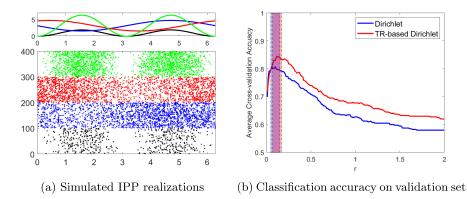
Fig 13: (a) Four intensity functions $\lambda_1(t)$ (black), $\lambda_2(t)$ (blue), $\lambda_3(t)$ (red), and $\lambda_4(t)$ (green) in the upper panel and 100 simulated processes with each of these functions in the low panel. (b) The average classification accuracy on validation set of the Dirichlet depth (blue) and TR-based Dirichlet depth (red), as a function of the weight coefficient r, and the optimal value/ranges of r (shaded area).

Figure 13(a) shows the simulated realizations and its intensity functions of the four Poisson processes we used in this example. By setting $K = 5$, i.e.,

5-Fold cross-validation method, we report the average classification on the validation set over five runs in Figure 13(b). As we can see from the plot, the optimal value(range) for Dirichlet depth method is $0.1([0.05, 0.15])$ with average accuracy of 0.8075, and for TR-based Dirichlet depth is $0.12([0.07, 0.17])$ with average accuracy of 0.845. In practice, the optimal r decided by the cross-validation method might not be unique, but many methods can help us to pick one from the optimal values.