The Annals of Applied Statistics
2017, Vol. 11, No. 2, 992–1010
DOI: 10.1214/17-AOAS1030

GENERALIZED MAHALANOBIS DEPTH IN POINT PROCESS AND ITS APPLICATION IN NEURAL CODING

BY SHUYI LIU AND WEI WU

Florida State University

In this paper, we propose to generalize the notion of depth in temporal point process observations. The new depth is defined as a weighted product of two probability terms: (1) the number of events in each process, and (2) the center-outward ranking on the event times conditioned on the number of events. In this study, we adopt the Poisson distribution for the first term and the Mahalanobis depth for the second term. We propose an efficient bootstrapping approach to estimate parameters in the defined depth. In the case of Poisson process, the observed events are order statistics where the parameters can be estimated robustly with respect to sample size. We demonstrate the use of the new depth by ranking realizations from a Poisson process. We also test the new method in classification problems using simulations as well as real neural spike train data. It is found that the new framework provides more accurate and robust classifications as compared to commonly used likelihood methods.

1. Introduction. In this paper, we introduce the notion of "depth" for point process data and apply it to neural coding problems in computational neuroscience. Point process Ross (1983) is a random collection of points falling in some space. In most applications, each point represents the time or location of an event, for example, lightning strikes over a period of time or earthquakes across a region. The most commonly used point process is the temporal Poisson process. Neural spike trains are time-dependent firing activities from neurons, which can be naturally treated as random realizations from a temporal point process. Parametric and semiparametric point processes have been the dominant models for neural spike trains. Many important and useful tools have been built for appropriate representations and efficient inferences in various neural systems [Box, Hunter and Hunter (1978), Brown et al. (2001), Kass, Ventura and Brown (2005)].

A point process model can be used to measure the probabilistic likelihood of any given spike train. However, such method only focuses on representations at each given time and, therefore, has a limited use in evaluating nonparametric statistics in the space of spike trains. Given a set of spike train samples, one may naturally pose questions such as: (1) "What is the central tendency or a representative template of the sample?" or more ambitiously, (2) "What is a center-outward ranking

Received December 2016; revised February 2017.
Key words and phrases. Point process, Mahalanobis depth, Poisson process, neural coding, spike train.

on the sample?" These questions are fundamentally important for nonparametric analysis on point process data, but are still underexplored in the statistics literature. The first question was partially addressed by recent investigations on the "average" of neural spike trains by Diez, Schoenberg and Woody (2012), Julienne and Houghton (2013), Mateu et al. (2015), Wu and Srivastava (2011). A comparison of these averages was recently given in Wesolowski, Contreras and Wu (2015).

Our goal in this paper is to address both questions. We will particularly focus on Question 2 because when a ranking is known, the sample point process observation with the highest rank represents the median, a natural choice of central tendency in the data. When the observations are ranked in a center-outward manner, the central one would be a template and those on borderlines could be outliers. This is actually the classical approach with regards to "depth" in multivariate data, where the value of depth represents how close a given observation is from the center. The depth has been extensively studied in nonparametric statistics, where various forms of depth have been developed to provide an center-outward ranking, such as location depth as in Tukey (1975), simplicial depth in Liu (1990), Mahalanobis depth in Liu and Singh (1993), Zonoid depth in Dyckerhoff, Koshevoy and Mosler (1996), projection depth in Zuo and Serfling (2000), expected convex hull depth in Cascos (2007) and geometrical depth in Dyckerhoff and Mosler (2011). The notion of depth was recently extended to functional data in López-Pintado and Romo (2009), Mosler and Polyakova (2012).

However, the depth in multivariate statistics is for a given dimension. Therefore, the above definitions cannot be directly applied to point process data because the number of events in each process observation varies. To the best of our knowledge, depth in the point process observations is still an under-explored area. In this paper, we propose to introduce the notion of depth for point process data, which generalizes depth in the multivariate case. This is a challenging task since there are two types of randomnesses in point process realizations: (i) the number of events, and (ii) the distribution of these events given that the number of events is known. We represent the depth of an observation s based on:

1. the probability of having $|s|$ (the cardinality of s) events, and
2. the conditional depth of s given $|s|$.

We propose the depth of a point process as a weighted product of the above two terms. Therefore, a typical point process realization should satisfy two important properties: (1) the number of events in the process has reasonable likelihood, and (2) the distribution of these events is typical or representative.

In this paper, we will provide a model to characterize both the number of events and their distribution in a point process. The latter one is certainly more challenging. In principle, when the number of events is known, any multivariate depth can be adopted to characterize the conditional depth. As an exploratory study on the ranking for point process, we propose to use a classical Mahalanobis depth for the conditional depth. The Mahalanobis-depth has some apparent benefits: (1) it is the

only depth based on mean and covariance in the multivariate data, (2) it provides an effective framework to describe central tendency as well as co-variability in the data, and (3) it properly ranks each observation with respect to the center based on a Gaussian kernel measurement.

This new notion of depth is desirable for data-driven studies on point process data. To build an effective depth, we need to answer several important questions:

- *Definitions*: What are the proper definitions of conditional mean and covariance in the point process data?
- *Computations*: How to estimate the conditional mean and covariance for a given set of observations?
- *Robustness*: If the estimated covariance is singular, how can we still effectively compute the depth value?
- *Simplification*: Is there any simplification on the representation and estimation for the commonly used Poisson process model?

In this paper, we will address each of these challenges. At first, we provide a definition on conditional mean and covariance. This is based on a recent work on Euclidean metric and summary statistics in point process [Wesolowski, Contreras and Wu (2015)]. A bootstrapping procedure is then proposed to estimate these summary statistics. If the estimated covariance matrix is singular, we provide a principal-component-based approach to robustly estimate the conditional depth value. Finally, we demonstrate that when the observations are from Poisson process, the precision matrix (inverse of the covariance) is tri-diagonal. The summary statistics can even have closed-forms when the Poisson process is homogeneous.

We emphasize that the conventional likelihood method Drazek (2013) cannot be used for the definition of the depth. We here point out the problems by using the following example: Suppose we get a realization of a sequence of events $x_1, \ldots, x_N$ from a Poisson process on the interval $[0, 1]$ with intensity function $\lambda(t)$ and total intensity $\Lambda = \int_0^1 \lambda(t)\, dt$. Then the likelihood of this sequence is

$$L(x_1, \ldots, x_N) = e^{-\Lambda} \prod_{i=1}^{N} \lambda(x_i). \tag{1.1}$$

We point out that this likelihood would be inappropriate to describe the *representiveness* of an observation in the Poisson process for the following reasons: 1. The likelihood will change w.r.t. the number of the events N. If $\lambda > 1$, then more events would have larger likelihood, but that does not imply the observation is more representative. 2. Template realization and outlier realization may have the same likelihood. This is true, for example, when the process is homogeneous where evenly distributed events (more representative) and any distribution with the same number of events will have the same likelihood. 3. The likelihood is sensitive to the noise. See Figure 1 for an example. A Poisson process has intensity function $\lambda(t) = 8 \cdot 1_{\{0.25 \le t \le 0.75\}}$, where $1_{\{\cdot\}}$ is an indicator function. We can easily see

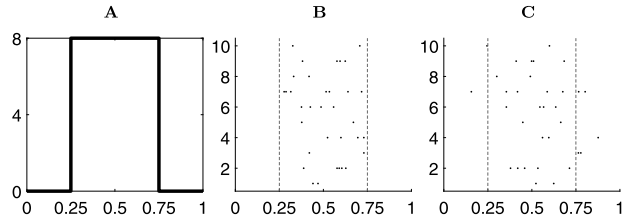

FIG. 1. *Nonrobustness of the likelihood method with respect to noise.* (A) *Intensity function of a Poisson process.* (B) 10 *simulated point processes. Each horizontal line indicates a process.* (C) *The simulated point processes with a Gaussian noise added at each event.*

that when a realization has noise out of the range $[0.25, 0.75]$, the likelihood in equation (1.1) will become 0.

In this paper, we introduce the notion of depth in point process data and propose a novel framework to define and compute this depth. The rest of the manuscript is organized as follows: In Section 2, we elaborate on the definition of the depth and provide computational procedures to effectively estimate it. We also study the asymptotics of the estimated depth with respect to sample size. In Section 3, we provide a thorough analysis for data from a Poisson process. The new method is illustrated with various simulations in Section 4. We then demonstrate its application using two neural spike train datasets. Finally, we discuss and summarize the work in Section 5.

2. Methods. We will at first provide the definition of the generalized Mahalanobis depth for point process. This definition is based on conditional mean and covariance in the data. We will then propose an algorithm to estimate these statistical terms.

2.1. *Definition of the generalized Mahalanobis depth.* Let S denote the set of all point processes in the time domain $[0, 1]$. We use $[0, 1]$ to simplify the notion. The method applies to any other finite interval. For any $s \in S$, its cardinality $|s|$ is a nonnegative integer and s can be treated as a vector in $[0, 1]^{|s|}$. Let $E = \bigcup_{d=0}^{\infty} [0, 1]^d$. The depth on the point process is a function $D : E \to \mathbb{R}^+, s \to D(s)$.

We represent the depth of a point process s based on: (1) the probability of having $|s|$ events, and (2) the conditional center-outward ranking $D(s \mid |s|)$. In this framework, we take the distribution of $|s|$ as Poisson and the conditional depth $D(s \mid |s|)$ using the Mahalanobis depth, that is, our definition of the depth function $D(\cdot)$ is given as follows.

DEFINITION 1. Given a point process s, its generalized Mahalanobis depth is defined as a weighted product:

$$D(s) := P\big(|s|\big)^r D\big(s \mid |s|\big),$$

with weight power $r > 0$. Here, the first term in the product is a normalized Poisson mass function, that is,

$$P(|s|) = \frac{\Lambda^{|s|}/|s|!}{\Lambda^{\lfloor\Lambda\rfloor}/\lfloor\Lambda\rfloor!},$$

where $\Lambda > 0$ is the total intensity and $\lfloor\cdot\rfloor$ denotes the floor function. For $|s| > 0$,

$$D(s \mid |s|) = \frac{1}{1 + (s - \mu_{|s|})^T \Sigma_{|s|}^{-1} (s - \mu_{|s|})},$$

is the Mahalanobis depth with mean $\mu_{|s|} \in \mathbb{R}^{|s|}$ and covariance $\Sigma_{|s|} \in \mathbb{R}^{|s|\times|s|}$. For $|s| = 0$, we simply let $D(s \mid |s|) = 1$.

In Definition 1, both the normalized Poisson term, $P(|s|)$ and Mahalanobis term, $D(s \mid |s|)$, are positive and can reach maximal value at 1 with $|s| = \lfloor\Lambda\rfloor$ and $s = \mu_{|s|}$, respectively. Therefore, $D(s) \in [0, 1]$ for any $s \in E$. The parameter $r > 0$ is a weight power for the normalized Poisson term. A larger r indicates more importance on the number of events for the depth on s.

The depth of a point process s depends on three parameters: Λ, $\mu_{|s|}$ and $\Sigma_{|s|}$. In general, these parameters are unknown and should be estimated from given observations.

The scalar parameter Λ is easy to estimate with the classical maximum likelihood method: Given a collection of independent point process observations $s_1, s_2, \ldots, s_K \in S$, the estimate $\hat{\Lambda}$ is $\frac{1}{K}\sum_{k=1}^{K} |s_k|$. We will study the estimation of $\mu_{|s|}$ and $\Sigma_{|s|}$ in Section 2.2.

According to the definition, $D(s \mid |s|)$ describes the depth of a point process observation when the number of events is fixed. This conditional depth corresponds to the classical Mahalanobis depth and has all important properties in a multivariate depth [Liu (1990), Zuo and Serfling (2000)] such as time-shift invariant, linear invariant, monotone on rays, upper semicontinuous, etc.

2.2. *Estimation of conditional mean and covariance.* In this section, we provide estimation details for the conditional mean $\mu_{|s|}$ and conditional covariance $\Sigma_{|s|}$. We at first review the sample statistics in point process realizations.

2.2.1. *Definitions.* Sample mean in point process has been defined for a set of realizations with the generalized Victor–Purpura metric in Wesolowski, Contreras and Wu (2015), Wu and Srivastava (2011), that is, for point processes $s_1, s_2, \ldots, s_K \in S$ in the time interval $[0, 1]$, their sample mean is defined as

$$\mu = \operatorname*{argmin}_{s\in S} \sum_{k=1}^{K} D_{\mathrm{GVP}}[\alpha](s_k, s)^2,$$

where $D_{\mathrm{GVP}}[\cdot]$ is the generalized Victor–Purpura metric between two processes and $\alpha > 0$ is the weight coefficient for matching event times between two processes.

In general, the defined mean does not have a closed-form because different point processes can have different numbers of events. An efficient and convergent algorithm has been proposed to conduct the estimation [Wesolowski, Contreras and Wu (2015)]. If all processes $s_1, s_2, \ldots, s_K$ have same number of events, then (for α sufficiently small) the sample mean has a conventional closed-form average, that is, if all processes $s_1, s_2, \ldots, s_K \in S_d$, where S_d denotes the set of all point processes on $[0, 1]$ with $d(> 0)$ events, their sample mean is given in the following form:

$$\hat{\mu}_d = \frac{1}{K} \sum_{k=1}^{K} s_k. \tag{2.1}$$

Based on this sample mean, the sample covariance can be defined as

$$\hat{\Sigma}_d = \frac{1}{K-1} \sum_{k=1}^{K} (s_k - \hat{\mu}_d)(s_k - \hat{\mu}_d)^T, \tag{2.2}$$

where ^{T} indicates transpose in matrix operation. It is apparent that the sample mean $\hat{\mu}_d$ and covariance $\hat{\Sigma}_d$ are consistent estimators for parameters μ_d and Σ_d, respectively.

Once conditional mean and covariance are estimated, we can estimate the conditional depth of any process s given $|s|$ in the following form:

$$\hat{D}\big(s \mid |s|\big) = \frac{1}{1 + (s - \hat{\mu}_{|s|})^T \hat{\Sigma}_{|s|}^{-1} (s - \hat{\mu}_{|s|})}. \tag{2.3}$$

Finally, the depth of the process s is computed as

$$\hat{D}(s) := \left(\frac{\hat{\Lambda}^{|s|}/|s|!}{\hat{\Lambda}^{\lfloor \hat{\Lambda} \rfloor}/\lfloor \hat{\Lambda} \rfloor!} \right)^r \hat{D}\big(s \mid |s|\big), \tag{2.4}$$

where $\hat{\Lambda} = \frac{1}{K} \sum_{k=1}^{K} |s_k|$ is the estimated total intensity.

2.2.2. *Bootstrapping estimation*. The above closed-forms on mean and covariance are certainly desirable. However, it is impractical to assume that all point process realizations have the same number of events in a given data set. We propose a bootstrapping approach to address this issue.

Given point process realizations $s_1, s_2, \ldots, s_K$, our goal is to compute conditional mean $\mu_d \in \mathbb{R}^d$ and covariance $\Sigma_d \in \mathbb{R}^{d \times d}$. Each process s_i is a vector in $\mathbb{R}^{|s_i|}$, $i = 1, \ldots, K$. In general, these vectors do not have the same length, so it is not possible to directly apply the average formula in equation (2.1). To solve this problem, we propose to apply a bootstrap method [Efron (1979)] to resample each

Algorithm 1 Compute conditional mean and covariance

Require: Given a sequence of point processes $s_1, s_2, \ldots, s_K$
  for $d = 1$ to N (N is pre-determined as $\max\{|s_i|\}$) **do**
    for $k = 1$ to K **do**
      If $|s_k| > d$, then uniformly randomly delete $|s_k| - d$ events in s_k.
      Otherwise, add $d - |s_k|$ events to s_k by uniformly resampling from s_k.
    end for
    Denote the K resampled realizations as $s_{1|d}, s_{2|d}, \ldots s_{K|d}$. Then

$$\hat{\mu}_d = \frac{1}{K}\sum_{k=1}^{K} s_{k|d},$$

$$\hat{\Sigma}_d = \frac{1}{K-1}\sum_{k=1}^{K}(s_{k|d} - \hat{\mu}_d)(s_{k|d} - \hat{\mu}_d)^T.$$

  end for
    return $\{\hat{\mu}_d\}_{d=1}^N, \{\hat{\Sigma}_d\}_{d=1}^N$

process s_i such that the resampled process $s_{i|d}$ has d events. Then the conditional mean μ_d and covariance Σ_d can be efficiently estimated. The detailed steps are given in Algorithm 1.

In the given set, the number of events in each process is up to $\max\{|s_i|\}$. Therefore, we let d vary from 1 to $\max\{|s_i|\}$ in our estimation (although, in principle, the algorithm can be used for any $d > 0$). One can also see the precision matrix $\hat{\Sigma}_{|s|}^{-1}$ (not the covariance $\hat{\Sigma}_{|s|}$) is actually needed in the depth computation. Therefore, we should work on the precision matrix directly if it has more convenient or efficient form. We will emphasize this point on the classical Poisson processes in Section 3.

2.2.3. *Depth estimation for singular covariance.* We have provided a computational approach in Section 2.2.2 to estimate the conditional mean and covariance. However, due to limited sample size the estimated covariance matrix can be singular and, therefore, the conditional depth, $\hat{G}(s \mid |s|)$, in equation (2.3) may not be directly obtained. We propose a solution to this problem in this subsection.

In general, let $x_1, \ldots, x_m \in \mathbb{R}^n$ be a set of i.i.d. multivariate normal random vectors with mean $\mu \in \mathbb{R}^n$ and singular covariance $\Sigma \in \mathbb{R}^{n\times n}$ of rank $k < n$. The likelihood cannot be estimated by the normal density function because the covariance is not invertible. This is a classical problem and can be addressed by a dimensional reduction approach as follows [Berrar, Dubitzky and Granzow (2009)]: Let $\Sigma = USU^T$ be the Singular Value Decomposition (SVD) of the covariance, where $U \in \mathbb{R}^{n\times n}$ is an orthogonal matrix and $S = \operatorname{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0) \in \mathbb{R}^{n\times n}$ is a diagonal matrix with k nonzero entries $\sigma_1, \ldots, \sigma_k$. Then $U^T x_1, \ldots, U^T x_m$ are i.i.d. normal random vectors with mean $U^T\mu$ and singular covariance S. For any vector X, we use $X(k_1 : k_2)$ to denote a vector with the k_1th entry to the k_2th entry in

Algorithm 2 Compute conditional depth when the covariance is singular

Require: Point process s with p number of events, conditional mean $\mu_p \in \mathbb{R}^p$, and conditional covariance matrix $\Sigma_p \in \mathbb{R}^{p\times p}$ with rank $d < p$.

1: Apply SVD on Σ_p such that $\Sigma_p = U\Sigma^* U^T$, where U is orthogonal and Σ^* is a diagonal matrix with first d diagonal entries being nonzero.

2: Let $U_d \in \mathbb{R}^{p\times d}$ be the matrix of U's first d columns.

3: The conditional depth can be computed as

$$D(s \mid |s| = p) = \frac{1}{1 + (U_d^T(s-\mu_p))^T(\Sigma^*)^{-1}(U_d^T(s-\mu_p))}.$$

return conditional depth $D(s \mid |s| = p)$

X. It is easy to see that $(U^T x_1)(1:k), \ldots, (U^T x_m)(1:k)$ are i.i.d. normal random vectors with mean $(U^T\mu)(1:k)$ and nonsingular covariance $\operatorname{diag}(\sigma_1, \ldots, \sigma_k)$, and $(U^T x_1)(k+1:n), \ldots, (U^T x_m)(k+1:n)$ are i.i.d. normal random vectors with mean $(U^T\mu)(k+1:n)$ and zero covariance. Note that, in practice, the random vectors for the $(k+1)$th to the nth entries may not be exactly constant (although the covariance is zero). Such variability can be treated as noise in finite sample and these entries should be excluded from the likelihood computation. Therefore, the likelihood of $x_1, \ldots, x_m \in \mathbb{R}^n$ can be computed using the likelihood of $(U^T x_1)(1:k), \ldots, (U^T x_m)(1:k)$ with the classical multivariate formula.

Algorithm 2 shows the detailed steps to compute the conditional depth for a given point process realization when the corresponding conditional covariance is not full rank.

2.3. *Consistency theory.* In this subsection, we study the asymptotic property of the estimated depth of a point process. For any $s \in S$, $|s|$ is a nonnegative integer and the depth on the point process s is the function $D(s)$ given in Definition 1.

In general, the parameters conditional mean $\mu_{|s|}$, conditional covariance $\Sigma_{|s|}$, and total intensity Λ in Definition 1 are unknown and need to be estimated from a training set of independent point process realizations $T_{\text{tr}} = \{s_1, s_2, \ldots, s_K\}$. Let $T_{|s|} = \{s_i \in T_{\text{tr}}, |s_i| = |s|, i = 1, \ldots, K\}$ denote all processes in T_{tr} that have $|s|$ events. We can estimate conditional mean $\mu_{|s|}$ and covariance $\Sigma_{|s|}$ using equations (2.1) and (2.2) based on this subset $T_{|s|}$. The estimated depth $\hat{D}(s)$ can then be obtained using equation (2.4), where $\hat{\Lambda} = \frac{1}{K}\sum_{k=1}^{K} |s_k|$ is the estimated total intensity. The following result shows that $\hat{D}(s)$ converges to $D(s)$ (a.s.) when the sample size K increases.

THEOREM 1. *Under the conditions given above, if* (1) *the covariance matrix* $\Sigma_{|s|}$ *is nonsingular, and* (2) $P(|s_i| = |s|) > 0$, *then* $\hat{D}(s)$ *converges to* $D(s)$ *almost surely with error bound*:

$$\left|\hat{D}(s) - D(s)\right| = O\left(\sqrt{\frac{\log\log K}{K}}\right).$$

The proof of this theorem follows standard procedure in large sample theory Ferguson (1996). We omit the details in this manuscript.

3. Depth in Poisson process. We have defined the notion of depth for point processes and provided algorithms for its estimation. It is well known that the Poisson process is the most classical point process. In this section, we will examine the proposed generalized Mahalanobis depth for a Poisson process. This study will illustrate the important properties of depth and provide insights for practical applications.

3.1. *Estimation in nonhomogeneous Poisson process.* In a Poisson process, given the number of events, d, the unordered event times are independent with each other, that is, let $\{x_i\}_{i=1}^d$ be unordered events times in a Poisson process realization with intensity function $\lambda(t)$ on $[0, 1]$. Then $\{x_i\}$ are a set of i.i.d. random variables with density function $f(t) = \frac{\lambda(t)}{\Lambda}$, where $\Lambda = \int_0^1 \lambda(t)\,dt$ [Karlin (1966)]. As a process is always ordered, the observation $s = (x_{(1)}, \ldots, x_{(d)})^T$ is simply the order statistics of $\{x_i\}_{i=1}^d$ and has the joint p.d.f. in the following form:

$$f_s(t_1, \ldots, t_d) = d! \left[\prod_{i=1}^{d} f(t_i)\right] \left[\prod_{i=1}^{d-1} 1_{t_i < t_{i+1}}\right].$$

Based on this p.d.f., the conditional mean and covariance are given as

$$\mu_d = E(s), \qquad \Sigma_d = E\big((s - \mu_d)(s - \mu_d)^T\big).$$

It is well known that the order statistics of independent realizations have Markovian property [Karlin (1966)]. In a Markovian chain, the past and future are independent given the present. Hence, the conditional precision matrix, Σ_d^{-1}, is tridiagonal [see Lauritzen (1996) for details]. This sparse structure assures the parameters in the covariance are essentially in the linear order of d, which can make the estimation of the precision matrix accurate and robust.

For a general nonhomogeneous Poisson process, there is no closed-form solution for the mean and precision matrix. However, once the intensity function is estimated, we can simulate more Poisson process realizations from it and then those statistical terms can be accurately estimated using a Monte Carlo method. Our computational approach is given in Algorithm 3.

3.2. *Estimation in homogeneous Poisson process.* Here, we study the estimation for a more special case—homogeneous Poisson Process. It turns out that the population conditional mean and covariance have closed-form estimates [Moghadam and Pazira (2011)]. Let $\{x_i\}_{i=1}^d$ be unordered events times in a homogeneous Poisson process with constant intensity λ on $[0, 1]$. Then $\{x_i\}$ are a set

Algorithm 3 Compute conditional mean and precision matrix in a Poisson process

Require: Poisson process realizations $s_1, s_2, \ldots, s_K$

for $k = 1$ to K **do**

Adopt the truncated Gaussian kernel smoothing method Robert (1995) on s_k to get a smooth density function on [0, 1], denoted as f_k.

end for

Take cross sectional mean of $\tilde{\lambda} = \frac{1}{K}\sum_{k=1}^{K} f_k$.

for $d = 1$ to N (N is pre-determined as $\max\{|s_i|\}$) **do**

1. Adopt $\tilde{\lambda}$ as intensity function and simulate K Poisson process realizations with d number of events. Denote them as $s_{1|d}, s_{2|d}, \ldots, s_{K|d}$.

2. Then

$$\hat{\mu}_d = \frac{1}{K}\sum_{k=1}^{K} s_{k|d},$$

$$\hat{\Sigma}_d = \frac{1}{K-1}\sum_{k=1}^{K}(s_{k|d} - \hat{\mu}_d)(s_{k|d} - \hat{\mu}_d)^T.$$

3. Take inverse of $\hat{\Sigma}_d$ as C. To get a tri-diagonal precision matrix, adopt the middle three diagonals of C and let other elements equal zero. We denote this new matrix as $\hat{\Sigma}_d^{-1}$.

end for

return $\{\hat{\mu}_d\}_{d=1}^N, \{\hat{\Sigma}_d^{-1}\}_{d=1}^N$

of i.i.d. uniform random variables. The observed process $s = (x_{(1)}, \ldots, x_{(d)})^T$ has the joint p.d.f. in the following form:

$$f_s(t_1, \ldots, t_d) = d!\left[\prod_{i=1}^{d-1} 1_{t_i < t_{i+1}}\right].$$

Based on the result from order statistics, we have

$$E(x_{(i)}) = \frac{i}{d+1}, \qquad i = 1, \ldots, d, \tag{3.1}$$

and

$$\mathrm{Cov}(x_{(i)}, x_{(j)}) = \frac{i(d+1-j)}{(d+1)^2(d+2)}, \qquad 1 \le i \le j \le d. \tag{3.2}$$

Hence, the mean and precision matrix have the following closed-forms, respectively:

$$\mu_d = \left(\frac{1}{d+1}, \frac{2}{d+1}, \ldots, \frac{d}{d+1}\right)^T,$$

$$\Sigma_d^{-1} = (d+1)(d+2)\begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}.$$

4. Results. In this section, we will illustrate the use of depth in point processes with simulated and real experimental data. We will rank realizations from a homogeneous Poisson process in one simulation. We will also apply the depth value to classify non-Poisson processes in another simulation. In the real experiments, we will test the classification performance using depth in neural spike trains from motor cortex and geniculate ganglion.

4.1. *Ranking Poisson processes.* Given a set of realizations from a Poisson process, we can use the generalized Mahalanobis depth to rank the data. For example, we randomly generate 50 realizations from a homogeneous Poisson process with $\lambda = 3$ on $[0, 1]$. In each process, let p_k denote the probability that there are k events. As the process is homogeneous Poisson, $p_k = \frac{e^{-3}3^k}{k!}$. For example, $p_0 = 0.05$, $p_1 = 0.15$, $p_2 = 0.22$, $p_3 = 0.22$, $p_4 = 0.17$, $p_5 = 0.10$, $p_6 = 0.05$ and $\sum_{k=7}^{\infty} p_k = 0.04$. Therefore, realizations should most likely have 2 or 3 events, but other number of events are also possible.

Figure 2(A) shows the 50 simulated realizations and each row denotes one process. Figure 2(B) shows the first five (1st–5th) ranked and last five (46th–50th) ranked processes based on the depth values when the weight power $r = 1$. It is apparent that the 1st ranked process has 3 (approximately) evenly distributed events on $[0, 1]$, which characterizes the homogeneous nature of the process and is a proper template. The 2nd to 5th ranked processes all have typical 2 or 3 events, and the events are also approximately evenly distributed. In contrast, the 46th to 50th ranked processes have nontypical number of events ($\geq$6) and the distributions are highly uneven.

We also show the ranking result (first five and last five) in Figure 2(C) for $r = 0.1$. With less weight on the normalized Poisson term, the number of events in each process is less important in ranking. We find that the first three processes are still the same as those in Figure 2(B) (for $r = 1$), but the fourth and fifth ones only have 1 event each. Note that these two single events are among the center of the interval $[0, 1]$, which represents typical homogeneous poisson process with one event. Overall, results in Figure 2 show that the depth values do provide a reasonable ranking among all observations.

4.2. *Important role of conditional covariance in the generalized Mahalanobis depth.* In this simulation, we generate point process observations from two different models. These two models share the same conditional mean but have different

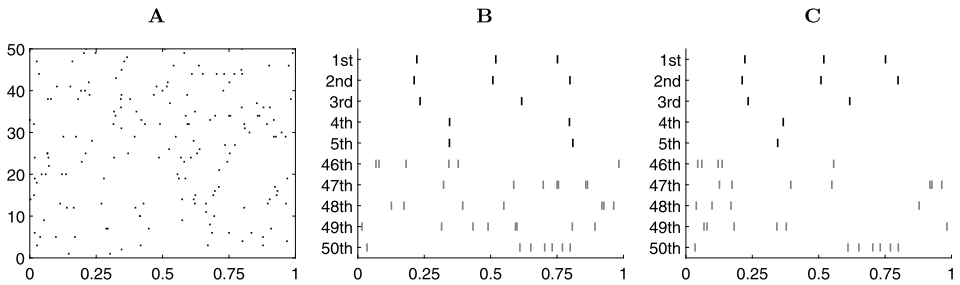

FIG. 2. (A) 50 *independent realizations from a homogenous Poisson process on* $[0, 1]$ *with* $\lambda = 3$. *Each row represents one realization.* (B) *The first and last* 5 *ranked realizations based on the depth value when the weight power* $r = 1$. (C) *Same as* B *except that* $r = 0.1$.

conditional covariances. We will test this difference by conducting classification task based on the proposed depth values and will illustrate the importance of covariance in the depth value definition.

100 point processes are simulated from each of the following two models:

- Model 1: generate $m(i) \sim \text{Poisson}(10)$, and then generate process $x_i = \{\frac{1}{m(i)+1} + \epsilon_i + \varepsilon_{1,i}, \frac{2}{m(i)+1} + \epsilon_i + \varepsilon_{2,i}, \ldots, \frac{m(i)}{m(i)+1} + \epsilon_i + \varepsilon_{m(i),i}\}, i = 1, \ldots, 100$.
- Model 2: generate $n(i) \sim \text{Poisson}(10)$, and then generate process $y_i = \{\frac{1}{n(i)+1} + \epsilon_{1,i} + \delta_{1,i}, \frac{2}{n(i)+1} + \epsilon_{2,i} + \delta_{2,i}, \ldots, \frac{n(i)}{n(i)+1} + \epsilon_{n(i),i} + \delta_{n(i),i}\}, i = 1, \ldots, 100$.

Here, $\epsilon_i \sim N(0, 0.1)$, $\epsilon_{j,i} \sim N(0, 0.1)$ and $\varepsilon_{j,i}, \delta_{j,i} \sim N(0, 0.05)$, $j = 1, 2, \ldots, n(i)$, $i = 1, \ldots, 100$. Note that in Model 1, same noise ϵ_i is used for each event. In contrast, different noises $\epsilon_{j,i}$ are used for events in Model 2. The simulated processes from the two models are shown in Figure 3(A) and (B), respectively. The processes look very similar across the two models.

Over the 100 point processes from each model, 50 of them are used as the training data and other 50 are used as the test data. We use the training samples to calculate sample conditional mean and covariance when the number of events varies from the minimum to the maximum in the training data by Algorithm 1. The estimated conditional means are shown in Figure 3(C) and (D) for the two models, respectively. We can see there is no significant difference on means between the two models for each conditional number of events. Figure 3(E) and (F) show the grayscale images of the covariance matrices conditioned on 10 events for both models. We can observe apparent differences. Basically, all covariance values in Model 1 are in the range 0.005 to 0.03, but values in Model 2 are in the range -0.01 to 0.025, that is, all event times in Model 1 are positively correlated, whereas some event times in Model 2 are negatively correlated.

Once conditional means and covariances are obtained, we can estimate the generalized Mahalanobis depth values of the 50 training processes in each model using equation (2.4). For each of the 100 testing processes (50 from each model),

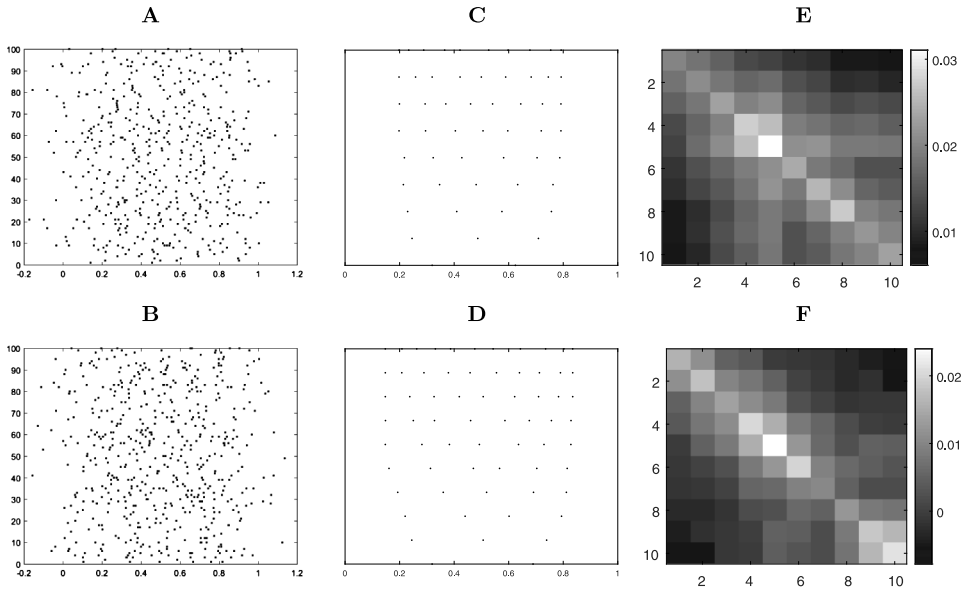

FIG. 3. (A) *Simulated* 100 *point processes from Model* 1. *Each row denotes one realization.* (B) *Same as* A *except for Model* 2. (C) *Estimated sample conditional means from Model* 1 *when the the number of events varies from* 2 *to* 10. *Each row denotes a mean under one condition.* (D) *Same as* C *except for Model* 2. *Here, the number of events varies from* 2 *to* 11. (E) *Grayscale image of the covariance matrix conditioned on* 10 *events in Model* 1. (F) *Same as* E *except for Model* 2.

we compute its depth value and then rank the value with respect to the 50 training values in each model. Finally, we classify this process to one of the two models in which the rank is relatively higher.

In the definition of generalized Mahalanobis depth, the conditional depth in equation (2.3) depends on conditional mean and covariance. To emphasize the importance of the covariance term, we can compute the mean-based depth by simply taking covariance as the identity matrix. Then the same ranking and classification procedure can be conducted on the testing processes. A third classification approach is the traditional likelihood method. Based on the Poisson process assumption, we can estimate the likelihood of each testing process and label it to the model in which the likelihood value is larger.

We compare the classification performance of the above three methods over all 100 testing processes in both models. The results are shown in Table 1. We can see that the classification performance using generalized Mahalanobis depth is superior to the other two methods: The averaged accuracy is 67% for the generalized Mahalanobis depth method, whereas the other two methods have around 47% to 59% accuracy. This indicates the proposed depth does provide useful quantification on the importance of each realization and the conditional covariances play a critical role in depth value formulation.

TABLE 1
Classification accuracy comparison

Method	Model 1	Model 2
Generalized Mahalanobis depth	**0.54**	**0.80**
Depth with mean only	0.54	0.64
Likelihood method	0.16	0.78

4.3. *Application in spike train data*. Here, we will apply the notion of depth in two real neuronal spike train datasets. Spike trains can be naturally treated as point process observations and we will use the proposed depth to conduct neural coding on the given spike trains.

4.3.1. *Motor cortical spike trains*. The first set of neural data was previously used in Wu and Srivastava (2011, 2013). Briefly, a microelectrode array was implanted in the arm area of primary motor cortex (MI) in a juvenile male macaque monkey (Macaca mulatta). Signals were filtered, amplified and recorded digitally using a Cerebus acquisition system (Cyberkinetics Inc.). Single units were manually extracted using Offline Sorter (Plexon Inc.). The subject was trained to perform a closed Squared-Path (SP) task by moving a cursor to targets via contralateral arm movements in the horizontal plane. Each sequence of 5 targets defined a path, and there were four different paths in the SP task (depending on the starting point). In this experiment, we recorded 60 trials for each path, and the total number of trials was 240. The recording time was normalized to 5 seconds. Figure 4 shows 5 example spike trains for each path.

Based on the proposed framework, we can compute the conditional means and covariances in each path. Figure 5 shows the estimated conditional means for four

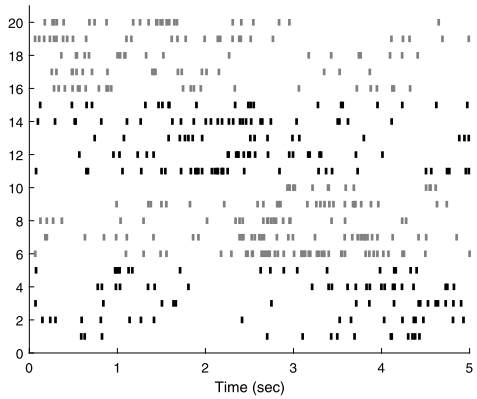

FIG. 4. 5 *example neural spike trains in each path*: *Trains* 1–5 *are for Path* 1, *trains* 6–10 *are for Path* 2, *trains* 11–15 *are for Path* 3 *and trains* 16–20 *are for Path* 4.

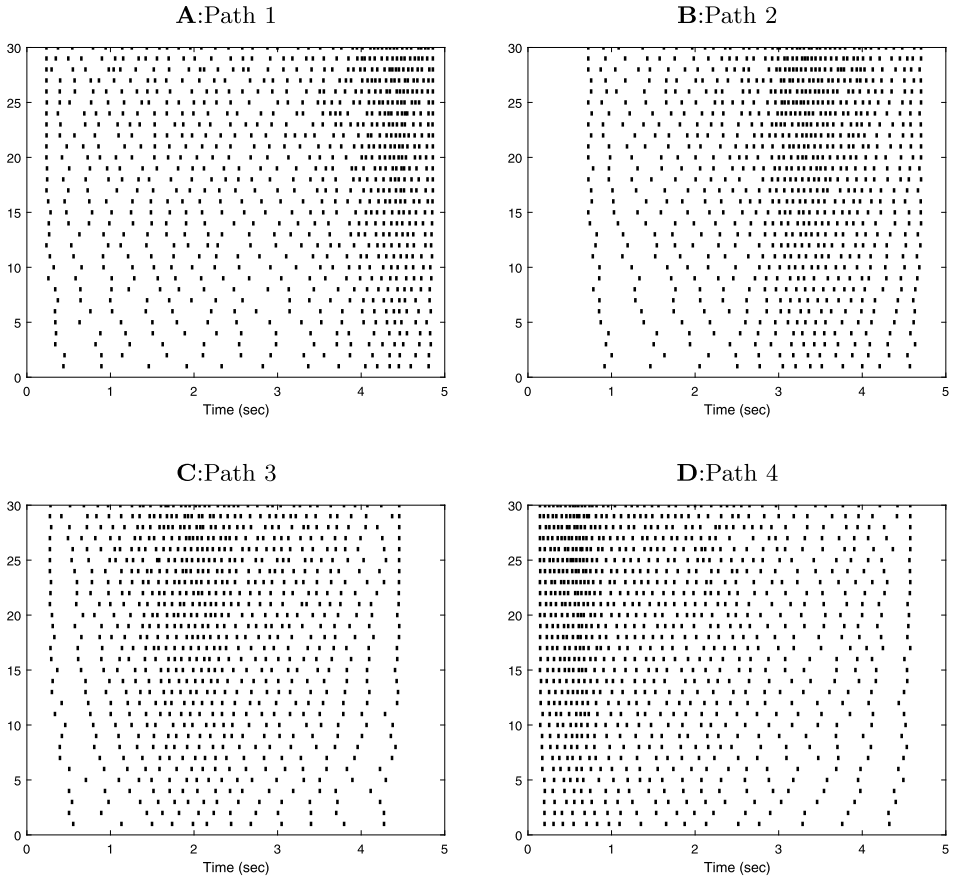

FIG. 5. 30 *conditional means in each path*. (A) *means conditioned on* 14 *to* 43 *spikes in path* 1. (B) *means conditioned on* 13 *to* 42 *spikes in path* 2. (C) *means conditioned on* 12 *to* 41 *spikes in path* 3. (D) *means conditioned on* 17 *to* 46 *spikes in path* 4.

paths using Algorithm 1. The difference among different paths is visually apparent for each conditional number of events. As compared to the example trains in Figure 4, these means properly represent the typical pattern in each path. Analogously to the simulation example, we compare the classification performance using the generalized Mahalanobis depth, depth with mean only and classical likelihood method. The results are shown in Table 2, where the classification rate of the generalized Mahalanobis depth method is 0.87, that of the depth with mean only is 0.75 and that of the likelihood method is 0.73. These results show that the proposed generalized Mahalanobis depth still has superior classification. We emphasize that our classification performance is robust to the parameter r, in a wide range of $[0.1, 100]$.

TABLE 2
Comparison of classification performance

Method	Classification accuracy
Generalized Mahalanobis depth	**0.87**
Depth with mean only	0.75
Likelihood method	0.73

4.3.2. *Geniculate ganglion spike trains*. This spike train dataset of rat geniculate ganglion neurons was previously used in Lawhern et al. (2011). Briefly, adult male Sprague-Dawley rat's geniculate ganglion tongue neurons were stimulated with 6 different solutions: KCl (salty), CA (sour), NaCl (salty), QHCl (bitter), MSG (umami) and Sucr (sweet). Each stimulus was presented 10 times. Stimulus trials were divided into three time regions: a 2-second pre-stimulus period, a 2.5-second stimulus application period and a 2-second post-stimulus period. We take two typical neurons for illustration: one is an electrolyte generalist, and the other is an acid generalist. The neural spike train data with respect to 6 different tastes from these two neurons are shown in Figure 6.

For the purpose of classification, we take 5 spike trains in each taste stimulus as training data to calculate mean spike numbers, conditional means and conditional covariance matrices by Algorithm 2. For the other 5 spike trains in each taste, we conduct classification based on the depth values. The classification results over the aforementioned three methods on each cell are shown in Table 3.

We can see from the results that the proposed Generalized Mahalanobis depth still has the best accuracy compared with the other two methods in both neurons. These results indicate again that the proposed depth provides a proper framework for measuring importance of each spike train in the given dataset. However, we also note that the improvement of our proposed method over depth-with-mean-only is

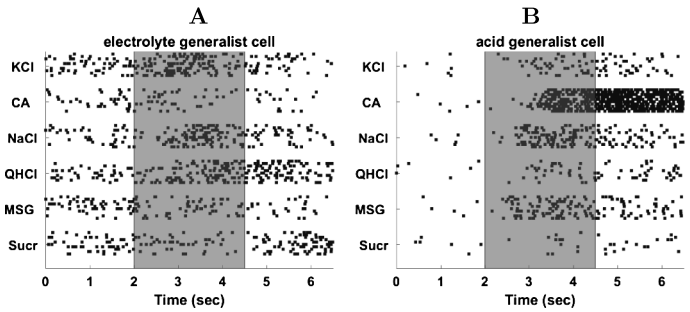

FIG. 6. *Single cell response by geniculate ganglion neurons to lingual stimulation with* 6 *basic tastes*. (A) *Neural spike trains in an electrolyte generalist* (*Cell* 1). (B) *Neural spike trains in an acid generalist* (*Cell* 2).

TABLE 3
Classification performance

Method	Cell 1	Cell 2
Generalized Mahalanobis depth	**0.70**	**0.76**
Depth with mean only	0.70	0.70
Likelihood method	0.47	0.33

marginal. This is expected as the number of training spike trains in each taste stimulation is only 5. With such an extremely small sample size, the covariance would not play an important role in neural decoding.

5. Summary. We have proposed a novel generalized Mahalanobis depth for point process observations. To the best of our knowledge, this is the first measurement of depth in point processes. The new framework extends the notion of depth on multivariate case, and it has three main advantages: (1) We can use depth to rank point process observations and identify the template and variability. We have shown this point in the Results section, where we rank the point process data by the values of depth. (2) For the Poisson process, the precision matrix is tri-diagonal and the estimation can be robust with respect to sample size. In the homogeneous case, we have the closed-forms to calculate the conditional mean and conditional precision matrix. (3) The new framework provides a more accurate classification than the depth-with-mean-only method and classical likelihood method in one simulation and two neural spike train data sets. This indicates the effectiveness of the new depth in practical application and importance of the covariance term.

We have used relative rank to classify point process observations. In the future, we will apply the same approach to conduct outlier detection. Based on the fact that depth is a statistical measure, we can plot depth control charts to observe pattern variation in the observed point processes. We will also explore the clustering of point process observations using the estimated summary statistics (i.e., conditional mean and covaraince). Finally, the conditional mean in this paper is given in the form of "mean" spike train with a generalized Victor–Purpura metric [Wesolowski, Contreras and Wu (2015)]. In principle, other spike train templates such as "consensus" in Victor and Purpura (1997), "prototype" in Mateu et al. (2015) and "average" in Julienne and Houghton (2013), can also be used. In the future, we will explore to adopt these templates in the definition of depth.

REFERENCES

BERRAR, D. P., DUBITZKY, W. and GRANZOW, M. (2009). *A Practical Approach to Microarray Data Analysis*. Springer, Berlin.

BOX, G. E. P., HUNTER, W. G. and HUNTER, J. S. (1978). *Statistics for Experimenters*: *An Introduction to Design*, *Data Analysis*, *and Model Building*. Wiley, New York. MR0483116

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 Comput*. **14** 325–346.

CASCOS, I. (2007). The expected convex hull trimmed regions of a sample. *Comput. Statist*. **22** 557–569. MR2358426

DIEZ, D. M., SCHOENBERG, F. P. and WOODY, C. D. (2012). Algorithms for computing spike time distance and point process prototypes with application to feline neuronal responses to acoustic stimuli. *J. Neurosci. Methods* **203** 186–192.

DRAZEK, L. C. (2013). Intensity estimation for Poisson Processes. The University of Leeds, School of Mathematics.

DYCKERHOFF, R., MOSLER, K. and KOSHEVOY, G. (1996). Zonoid data depth: Theory and computation. In *OMPSTAT*: *Proceedings in Computational Statistics* 12*th Symposium held in Barcelona, Spain*, 1996 235–240. Physica-Verlag, Heidelberg.

DYCKERHOFF, R. and MOSLER, K. (2011). Weighted-mean trimming of multivariate data. *J. Multivariate Anal*. **102** 405–421. MR2755006

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist*. **7** 1–26. MR0515681

FERGUSON, T. S. (1996). *A Course in Large Sample Theory*. Chapman & Hall, London. MR1699953

JULIENNE, H. and HOUGHTON, C. (2013). A simple algorithm for averaging spike trains. *J. Math. Neurosci*. **3** Art. 3, 14. MR3042443

KARLIN, S. (1966). *A First Course in Stochastic Processes*. Academic Press, New York. MR0208657

KASS, R. E., VENTURA, V. and BROWN, E. N. (2005). Statistical issues in the analysis of neuronal data. *J. Neurophysiol*. **94** 8–25.

LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. The Clarendon Press, Oxford Univ. Press, New York. MR1419991

LAWHERN, V., NIKONOV, A. A., 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. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist*. **18** 405–414. MR1041400

LIU, R. Y. and SINGH, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc*. **88** 252–260. MR1212489

LÓPEZ-PINTADO, S. and ROMO, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc*. **104** 718–734. MR2541590

MATEU, J., SCHOENBERG, F. P., DIEZ, D. M., GONZÁLEZ, J. A. and LU, W. (2015). On measures of dissimilarity between point patterns: Classification based on prototypes and multidimensional scaling. *Biom. J*. **57** 340–358. MR3324443

MOGHADAM, S. A. and PAZIRA, H. (2011). The relations among the order statistics of uniform distribution. *Trends in Applied Science Research* **6** 719–723.

MOSLER, K. and POLYAKOVA, Y. (2012). General notions of depth for functional data. Available at arXiv:1208.1981.

ROBERT, C. (1995). Simulation of truncated normal variables. *Stat. Comput*. 121–125.

ROSS, S. M. (1983). *Stochastic Processes*. Wiley, New York. MR0683455

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

VICTOR, J. D. and PURPURA, K. P. (1997). Metric-space analysis of spike trains: Theory, algorithms and application. *Network* **8** 127–164.

WESOLOWSKI, S., CONTRERAS, R. J. and WU, W. (2015). A new framework for Euclidean summary statistics in the neural spike train space. *Ann. Appl. Stat*. **9** 1278–1297. MR3418723

WU, W. and SRIVASTAVA, A. (2011). An information-geometric framework for statistical inferences in the neural spike train space. *J. Comput. Neurosci.* **31** 725–748. MR2864743

WU, W. and SRIVASTAVA, A. (2013). Estimating summary statistics in the spike-train space. *J. Comput. Neurosci.* **34** 391–410. MR3061973

ZUO, Y. and SERFLING, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28** 461–482. MR1790005

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306-4330
USA
E-MAIL: s.liu@stat.fsu.edu
wwu@stat.fsu.edu