Annals of the Institute of Statistical Mathematics manuscript No.
(will be inserted by the editor)

On distances between point patterns and their applications

Jorge Mateu · Frederic P Schoenberg · David M Diez

Received: date / Revised: date

Abstract Various classes of distance metrics between two point process realizations observed on a common metric space are outlined. Examples include spike-time distance and its variants, proposed by Victor and Purpura (1997), and this list is extended to include cluster-based distances and distances based on classical statistical summaries of point patterns, such as the K-function or LISA functions. Applications to the summary and description of collections

Jorge Mateu

Department of Mathematics, University Jaume I, E-12071 Castellon, Spain

E-mail: mateu@mat.uji.es

Frederic P Schoenberg

Department of Statistics, UCLA, 8125 Math-Sciences, Los Angeles, CA 90095, USA

E-mail: frederic@stat.ucla.edu

David Diez

Department of Statistics, UCLA, 8125 Math-Sciences, Los Angeles, CA 90095, USA

E-mail: david.m.diez@gmail.com

of independent point process realizations via prototypes or multidimensional scaling are explored.

Keywords K-function · LISA functions · point processes · Poisson processes · prototypes · spike-time distance

1 Introduction

Distance methods are fundamental in almost every area of statistics. Indeed, many estimators such as regression or m-estimators are defined as minimizers of classical distance functions, and distance metrics play an obvious role in cluster analysis and multidimensional scaling, to name just a few of many examples.

However, distance methods have received relatively scant attention in the analysis of point processes to date. An exception is the work of Victor and Purpura (1997), who proposed several distance metrics, including spike-time distance, useful for describing neuronal spike trains. The list of distances in Victor and Purpura (1997) is far from exhaustive, however, and certain alternative types of distances might be more useful for clustered or inhomogeneous point processes or for point processes in high-dimensional spaces.

The purpose of this article is to outline several types of distances between two point process realizations, X and Y, each observed on the same metric space S. We also discuss some examples and applications of the proposed distance metrics to cluster analysis and prototype determination.

The paper is outlined as follows. In Section 2 we discuss spike-time distance and its variants which involve matching individual points of X to corresponding points in Y. Section 3 describes distances that might be especially useful for clustered point processes. Distances based on classical functional or scalar descriptors of point processes, including estimators of first and second moments of the processes, or classical test statistics based on these moments, are described in Sections 4 and 5. Applications to summaries of collections of independent realizations of a point process via prototypes or multidimensional scaling are given in Section 6.

2 Pointwise distances

One family of distances is characterized by a raw transformation of one pattern into another by actions on individual points. For instance, the **spike-time distance** metric developed by Victor and Purpura (1997) is the minimum cost associated with the transformation of one point pattern into another by deleting, adding, and moving points. The left panel of Fig. 1 represents a transformation of pattern X into Y via these actions. Those points in X not moved are deleted from X, and those points in Y to which no point in X corresponds are added to X. If we describe such a transformation as T, then the cost associated with T is

$$C(T|X,Y) = p_d|X_{delete}| + p_a|Y_{add}| + \sum_{x \in X_{move}} p_m d_x \qquad (1)$$

where p_a, p_d, and p_m are penalty parameters, X_{delete}, Y_{add}, and X_{move} are subsets of X and Y that are deleted, added, and moved, respectively, d_x is the distance a point x in X_{move} is moved, and $|S|$ represents the number of points in a set S. The spike-time distance, $d_s(X, Y)$, is defined as the infimum of equation (1) over all such possible transformations T. The transformation shown in the left panel of Fig. 1 is the transformation minimizing (1) for a particular set of parameters. Victor and Purpura (1997) showed that spike-time distance is a well-defined distance metric, and they also considered other closely related distances.

Another useful distance function that is also defined in terms of such transformations of X into Y is the **nearest-point distance**, where each point x in X is simply moved to its nearest neighbor in Y. For convenience, call this point y_x. Unlike spike-time distance, nearest-point distances are computed without allowing the addition or deletion of points. The nearest-point distance is defined simply as

$$d_n(X, Y) = \sum_{x \in X} ||x - y_x||$$

where $||\cdot||$ may represent Euclidean distance for point patterns in $\mathbf{R}^k$, or some other distance metric for point patterns in a more general metric space. Note that for two distinct points x, x' in X, we may have $y_x \equiv y_{x'}$, and that in general, $d_n(X, Y) \neq d_n(Y, X)$, so that d_n is not formally a distance metric. However, one may also consider the sum of these two distances as a symmetric nearest-point distance metric:

$$d_N(X, Y) = d_n(X, Y) + d_n(Y, X)$$

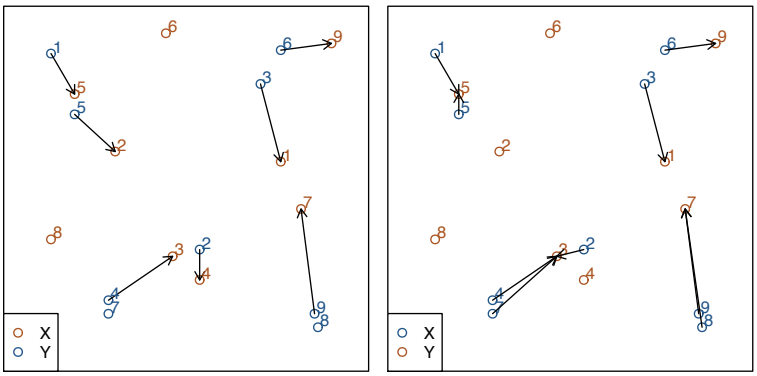

Fig. 1 Spike-time distance (left) and nearest-point distance for moving X to Y (right).

3 Distances for cluster processes

When a point process is highly clustered, metrics such as spike-time distance tend to yield prototypes that do not reflect a typical sample point pattern. Clustered process may thus require a separate family of distances where movements of collections of points are permitted. For example, the following metric may be called **cluster distance**. Let T represent a transformation of X into Y, sequentially moving collections of points in X. Thus, T is itself a sequence of transformations, t_i, where each t_i moves a subset X_i of points by a vector z_i (see the first two panels of Fig. 2). Then the cost of T may be defined as

$$p_d|X_{delete}| + p_a|Y_{add}| + \sum_i p_m|X_i|^q||z_i|| \tag{2}$$

where q is a parameter in $[0, 1)$. The cluster distance, $d_c(X, Y)$, is the infinum cost over all such transformations.

One may instead define a distance for clustered point processes via first aligning clusters in X with clusters in Y and subsequently applying simple spike-time distance. This adjustment is shown in the right panel of Fig. 2. For instance, let $\{R_j\}$ represent a set of disjoint and concave regions of the space that contain all the points of X. One may translate all of the points in a given region by some fixed vector and repeat for each region, assigning a cost of p_c per unit distance to each translated region, and then spike-time distance may be applied after these regions are moved. An illustration is given in Fig. 2. This declustered spike-time distance is defined as the minimum cost over all such transformations and choices of $\{R_j\}$.

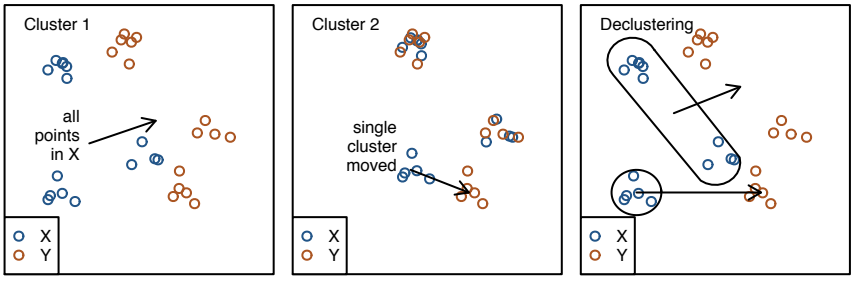

Fig. 2 The first two steps in the cluster distance are shown in the left and center panels. The right panel represents the process of removing the clustering characteristics prior to applying spike-time distance.

4 Distances based on functional summaries

4.1 Model-based distances

Given models for the point processes giving rise to the point patterns X and Y, one may define the distance between X and Y in terms of the differences between characteristics of these models. For instance, if the point processes X and Y are characterized by their conditional intensities $\lambda_X(x)$ and $\lambda_Y(x)$, respectively, then one measure of the difference in these point process models is the integral of the squared difference of the two conditional intensities over the observation region S:

$$\int_S (\lambda_X(x) - \lambda_Y(x))^2 dx. \tag{3}$$

This is illustrated in the left panel of Fig. 3, where the intensity functions have been estimated based on the data shown at the top of the panel. These methods readily extend to a variety of other summaries, such as the integrated squared difference between the overall mean, or second moment measure, or higher moments or cumulants of the processes. Of course, the conditional intensities in (3) may be replaced by their expected values, their overall intensities, or in the spatial point process setting, by the Papangelou intensities (see e.g. Daley and Vere-Jones (2003)).

4.2 Distances based on classical functional statistical summaries

Differences in point process behavior can also be characterized by comparing classical summary measures of the first or second moments for each of the point

patterns. For instance, given two point patterns X and Y, one could imagine looking at an estimate of the overall intensity of X, e.g. a kernel estimate, and a similar estimate of the intensity of Y. Alternatively, if each of the point patterns is one-dimensional, then one could look at the empirical cumulative distribution function for each point pattern as a statistical summary of the realization. A further alternative would be to take the estimated K-function or its derivative, the estimated reduced 2nd moment measure, for each pattern, and examine the difference.

As an illustration, one could compare the integrated squared difference between estimates of Ripley's K-function for two point patterns:

$$\int (K_1(r) - K_2(r))^2 dr \tag{4}$$

This distance is illustrated in the right panel of Fig. 3.

First and second order measures represent a valuable description of a spatial point process but do not generally uniquely characterize a point process. Additional important summary descriptions are given by distance methods based on measures of some distances between points. The *nearest neighbor distance* D may be defined as the distance from a point of the pattern to the closest of the other points in the same pattern. The *empty space distance* F is the distance from an arbitrary fixed location to the nearest point of the pattern. For a given point pattern, a set of distances (nearest neighbor or empty space distance) can be summarized by means of their corresponding distribution functions. Let $G(r)$ denote the nearest-neighbor distribution function, and $F(r)$ the corresponding first contact distribution function or empty space

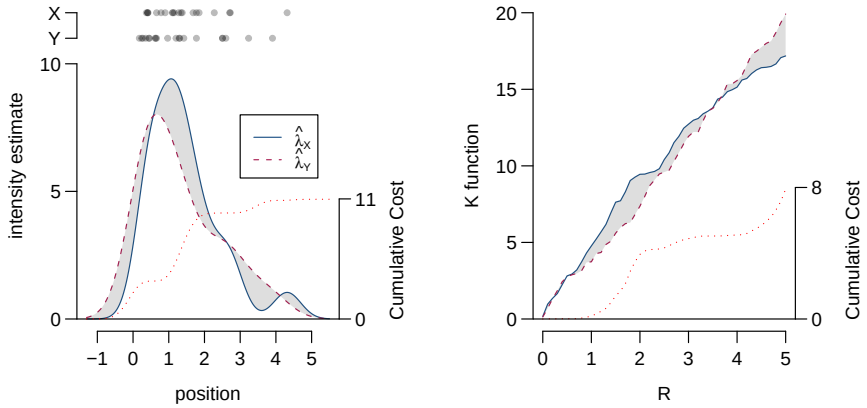

Fig. 3 Left: A comparison of two intensity models using Equation (3). In this illustration, the intensities λ_X and λ_Y are estimated by kernel smoothing the points in X and Y, respectively. Right: an example using Equation (4) to characterize the difference in clustering behavior of two patterns using Ripley's K function.

function. These distribution functions may be estimated from the observed point patterns X and Y using conventional methods.

In practice, each point pattern is typically observed in a bounded region S, and without precautions, these boundaries can lead to biased estimates. This problem is known in the context of spatial statistics as edge-effects. Different edge-corrected estimators of the functions F and G have been proposed (see e.g. Cressie (1993)). There is a clear analogy with censoring in the context of survival analysis Baddeley and Gill (1997). The different distances observed within a single point pattern are really censored distances. Let d_i denote the observed distance from the ith point of the point pattern to its nearest neighbor within the sampling window, W, and c_i its distance to the complement of the sampling window W. If $c_i < d_i$ the real nearest neighbor could be outside the

window and, in this case, the real and unknown nearest neighbor distance d_i fulfills $d_i > c_i$, and the observation is censored. A similar comment applies to empty space distances. In both cases the censoring distance is the distance from the sampling point to the frame window.

Given a statistical summary such as the estimated F or G-function of the point processes, the dissimilarity between point processes X and Y can be defined as the distance between the corresponding estimated functional summaries. For instance, let $\hat{F}_X$ and $\hat{F}_Y$ be the corresponding estimated empty-space distribution functions for X and Y. The dissimilarity between X and Y can thus be defined as

$$D(X,Y) = d(\hat{F}_X, \hat{F}_Y) \tag{5}$$

where d stands for a metric between the functions. For instance, one may use the L_2 metric

$$D_F^2(X,Y) = \int_0^{r_0} (\hat{F}_X(r) - \hat{F}_Y(r))^2 dr \tag{6}$$

or the L_∞ metric:

$$D_F^\infty(X,Y) = sup_r \|\hat{F}_X(r) - \hat{F}_Y(r)\| \tag{7}$$

By replacing the empty-space distribution function with the nearest-neighbor distribution function G or K-function K, one can similarly define $D_G^2(X,Y)$, $D_G^\infty(X,Y)$, $D_K^2(X,Y)$, and $D_K^\infty(X,Y)$, respectively. Note that sampling variability in estimates of $K(r)$ tend to increase with r and this can have a great influence on the value of the dissimilarity measure. This suggests the use of the L-function proposed by Besag (1977), a variance stabilized version of the K-function.

4.3 Distances based on LISA functions

An alternative class of functions one may use to characterize each process X and Y is based on LISA functions. Given point patterns $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_m\}$, one may calculate the set of LISA functions for each pattern corresponding to any individual or spatial location. Denote both sets by $\{l_i^X(r), i = 1, \ldots, n\}$ and $\{l_j^Y(r), j = 1, \ldots, m\}$. Then one may derive several possibilities for distances between X and Y, such as

(a) Define

$$\begin{aligned} A_L^X &= 1/(n \times (n-1)) \sum_i \sum_j d(l_i^X(r), l_j^X(r)) \\ &= 1/(n \times (n-1)) \sum_i \sum_j \int (l_i^X(r) - l_j^X(r))^2 dr \end{aligned} \tag{8}$$

as the average of all pairwise distances between LISA functions of points in X. Define the same average for points in Y, say, A_L^Y. Then, a similarity or dissimilarity measure is given by

$$d_L(X, Y) = A_L^X - A_L^Y. \tag{9}$$

(b) Define the average LISA function coming from the set of LISA functions for points in X and Y. Denote them by $l^X(r)$ and $l^Y(r)$. Note that $l^X(r)$ is a function itself. Then one may define a distance measure via

$$d_l(X, Y) = d(l^X(r), l^Y(r)) = \int (l^X(r) - l^Y(r))^2 dr. \tag{10}$$

4.4 The proximity function of an individual to a population based on LISA distances

Let $d_X(i,j) = d(l_i^X(r), l_j^X(r)), i,j = 1,\ldots,n$ the set of pairwise distances between LISA functions for points in X. Following Cuadras et al. (1997) we define the *geometric variability* for pattern X as

$$V_{d_X} = \frac{1}{2n^2}\sum_{i,j} d_X^2(i,j) \tag{11}$$

Then the *proximity function* for the $i-th$ point in X $i = 1,\ldots,n$ is given by

$$\Phi_X^2(i) = \frac{1}{n}\sum_{j=1}^{n} d_X^2(i,j) - V_{d_X} \tag{12}$$

Hence the value of the associated proximity-based density function for the $i-th$ point in X for $i = 1,\ldots,n$ is given by

$$f_X(i) = exp\{-\frac{1}{2}\Phi_X^2(i)\} \tag{13}$$

One may proceed similarly with pattern Y defining V_{d_Y}, $\Phi_Y^2(j)$ and $f_Y(j)$ for $j = 1,\ldots,m$.

Finally, the distance between patterns X and Y may be defined as a measure of similarity/dissimilarity between the densities f_X and f_Y, as $d_P(X,Y) = d(f_X, f_Y)$.

5 Distances based on numerical test statistics

One could alternatively compare point processes X and Y based on single numerical summaries of each point process. For instance, given the same parametric model for the two processes, one could obtain estimates θ_X and θ_Y of

the parameter vector governing each of the two processes, given the realizations X and Y, respectively. The distance between X and Y could then be defined simply as $||\theta_X - \theta_Y||$.

A related idea is to describe the similarity between X and Y in terms of a numerical summary of the degree of clustering or inhomogeneity in each. An example would be a generalization to the point process context of the log rank test (Mantel and Haenszel 1959) or Kolmogorov-Smirnov type statistic (Fleming and Harrington 1981), where the time or duration usual in the context of survival analysis is replaced here by the distance observed.

For instance, consider the following generalization of the log rank test of Mantel and Haenszel (1959). In survival analysis, the time instant where an event is observed is frequently called (observed) failure time. Here this phrase is replaced by (observed) failure distance, i.e., the distance between a point of the pattern and its nearest neighbor or the distance between a sampling point and the nearest event of the point pattern, provided that these distances are smaller than the distances to the boundary of the window. In order to construct the log rank test, a sequence of 2×2 tables is built over the distances (one for each failure distance, d_j), where the risk set at that distance is classified into a 2×2 table, according to group and event status. Let us present some notations: r_j stands for the size of the risk set in the interval $[d_{j-1}, d_j)$ and f_j the number of failures. The size of the risk and failure sets for each group can also be defined. In this way, for the first group, they are denoted by r_{1j} and f_{1j} respectively. Differences between the observed and the expected events

(conditioning on the margins of the 2×2 table) on the first group are added up over distances and squared in order to calculate the numerator of the log rank statistic:

$$U = \sum_j (f_{1j} - g_{1j}), \tag{14}$$

where g_{1j} stands for the expected number of failures in the first group, $g_{1j} = f_j \cdot r_{1j}/r_j$. The denominator is the sum of the variances of the number of events on the first group within each 2×2 table, which is obtained by using the hyper-geometric distribution:

$$Var(U) = \sum_j t_{1j} \cdot f_j/r_j(1 - f_j/r_j)(r_j - r_{1j})/(r_j - 1) \tag{15}$$

The null distribution of $U^2/Var(U)$ is approximately chi-squared with one degree of freedom. The log rank test is more sensitive to differences at the right tails of the two survival distributions and more suitable for detecting departures when hazards are proportional (Lee 1992). There are cases in which this test is not very effective, and one may use a Kolmogorov-Smirnov type statistic, i.e. the supremum of appropriately scaled empirical processes. This test is sensitive to differences that are large at a particular point. A detailed explanation can be found in Fleming and Harrington (1981).

6 Applying measures to a collection of patterns

Distance measures can be used in the identification of prototypical patterns or with multidimensional scaling for classification. Each technique is considered

in the context of a collection of patterns that come from either one or many point processes.

6.1 Prototypical patterns

One application of distance measures is in the construction of a **point pattern prototype**, which is the typical pattern of a collection. We identify the prototype $P_d(\mathcal{C})$ of the collection $\mathcal{C}$ based on the distance measure d as the pattern that minimizes the following loss function:

$$argmin_P \sum_{X_i \in \mathcal{C}} \alpha_i d(P, X_i)$$

where α_i is the weight corresponding to point pattern X_i. The weights α_i are typically chosen to be unity unless certain point patterns are deemed more important than others, or if the point patterns are measured with differential error.

Prototypes are useful for identifying the typical pattern in a collection, much like the median is used to describe the typical observation in a univariate data set. The prototype has generally been used to construct robust representations of a collection of patterns using spike-time distance as the distance metric (Schoenberg and Tranbarger 2008; Tranbarger 2005).

Kernel density and prototype summaries are provided for ten simulated patterns are shown in Fig. 4. The kernel density is a pooled estimate based on all patterns and is heavily influenced by pattern 10. The prototype, using spike-

time distance, is shown to be resistant to the tenth pattern and characterizes the typical pattern in the collection.

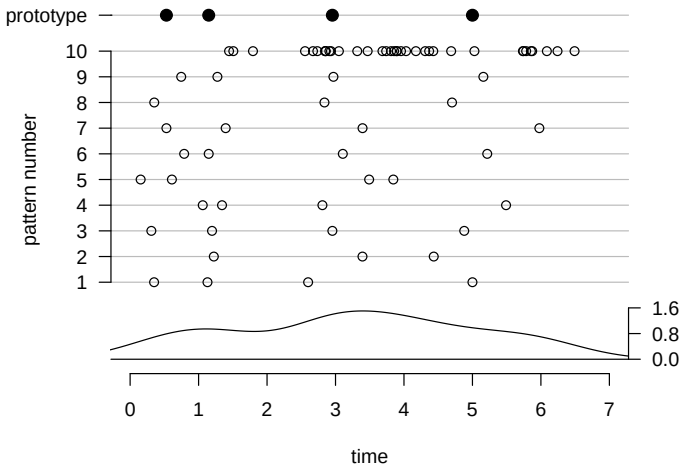

Fig. 4 Kernel density and prototype summaries for a collection of ten patterns. Each pattern is represented by a sequence of events at times denoted by circles.

A two-dimensional example of a prototype is shown in Fig. 5. Here the sample collection contains 10 point patterns, nine of which are realizations of an inhomogeneous Poisson process with an expected value of 15 points per point pattern, and the last point pattern is a realization of a homogeneous Poisson process with an expected value of 42 points per realization. Just as in the one-dimensional case, the prototype, shown in the bottom-right panel, is resistant to a single atypical pattern.

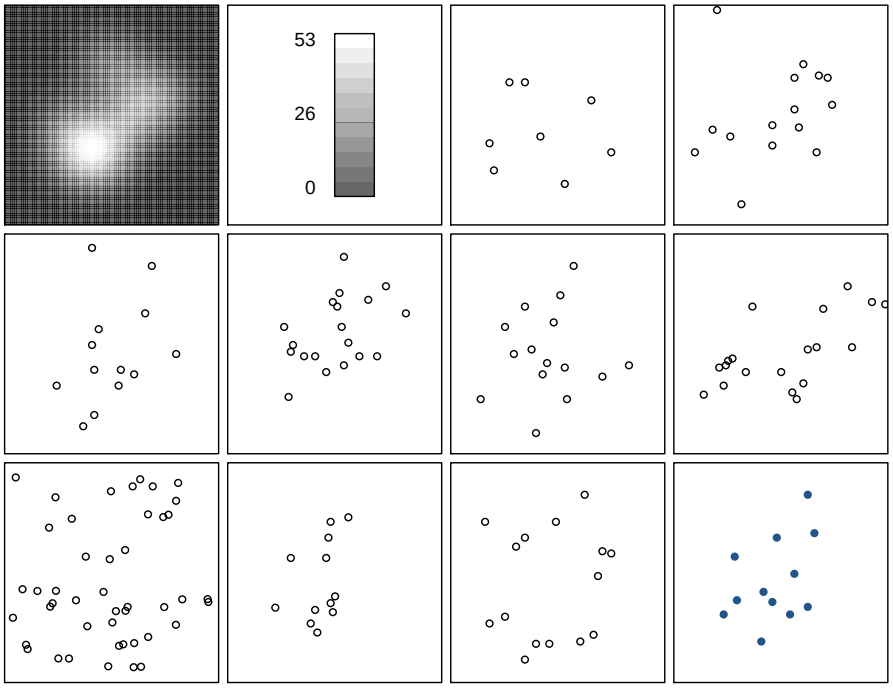

Fig. 5 The top-left panel shows an intensity function over the space $[0,1] \times [0,1]$. The second panel in the first row represents a legend of the intensity, and the next 9 panels show simulations of an inhomogeneous Poisson process based on the intensity, except the bottom-left panel, which is a realization of a homogeneous Poisson process with $\lambda = 42$. The prototype of these 9 point patterns, using spike-time distance with parameters chosen by default according to the recommendations in Diez et al. (2010), is shown in the bottom-right panel.

6.2 Multidimensional scaling

Classical multidimensional scaling (MDS) is useful for pattern classification based on a distance metric. Given a collection of point patterns, $\mathcal{C} = \{X_1, ..., X_n\}$, a distance matrix D is computed where $D_{i,j} = d(X_i, X_j)$ for some distance measure d. MDS uses this distance matrix to estimate respective locations of

the patterns in $\mathbf{R}^k$ (k selected by the user). Each pattern is itself represented by a point, and MDS embeds these points in locations of $\mathbf{R}^k$ such that the resulting Euclidean distances between the patterns, $\widetilde{D}$, approximates D by minimizing the following loss function:

$$L(D, \widetilde{D}) = \sum_{i \neq j} \left[D_{ij} - \widetilde{D}_{ij} \right]^2 .$$

See e.g. Venables and Ripley (2002) for additional details on classical MDS and its variants, such as Kruskal's MDS and Sammon mapping.

MDS can be useful in both identifying groupings of points – in this case, groups of patterns – and in classification. If classifications are provided on a training set, then a distance metric may be applied to compute the distance matrix D. MDS then may be applied to find respective locations of the patterns in $\mathbf{R}^k$, where classification techniques may be applied to classify a test data set.

The left panel of Fig. 6 shows 15 patterns in one-dimensional space. Each pattern arose from a Poisson process with an overall mean of 12 points per pattern. Patterns 1-5 arose from one intensity function and patterns 6-15 from a second intensity function. The right panel of Fig. 6 shows MDS applied to these fifteen patterns using spike-time distance. We see a grouping of the patterns based on their original intensity functions.

Multidimensional scaling may be applied to any of the distance metrics introduced. Fig. 7 shows a second example of 15 two-dimensional patterns and their representation as points in $[0, 1] \times [0, 1]$ where the distance D_K^2 was used to construct the dissimilarity matrix. Each point pattern is a realization

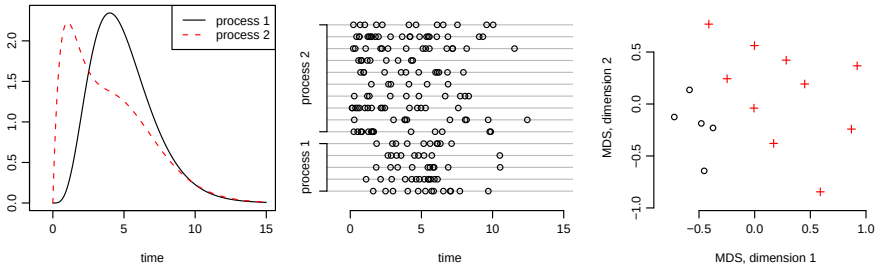

Fig. 6 Left: Two Poisson intensities, each with a total intensity of 12. Middle: 15 patterns, five from the first Poisson process and ten from a second Poisson process depicted in the left panel. Right: Spike time distance was used as the metric to construct the matrix D, and multidimensional scaling applied to all 15 patterns. The black circles denote the first five patterns and the red crosses mark those of the second process.

of one of two Neyman-Scott processes with different parameters. Again, one sees that MDS is quite successful at classifying the point patterns into two groups.

References

Baddeley, A. and Gill, R. D. (1997). Kaplan-meier estimators of distance distributions for spatial point processes. *Annals of Statistics*, 25(1):263–292.

Besag, J. (1977). *Contribution to the discussion of Dr Ripley's paper.* B 39: 193-195. J R Statistical Society.

Cressie, N. A. C. (1993). *Statistics for Spatial Data, revised ed.* New York: Wiley.

Cuadras, C. M., Atkinson, R. A. and Fortiana, J. (1997). Probability densities from distances and discrimination. *Statistics & Probability Letters*, 33:405–411.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Volume 1: Elementary Theory and Methods.* New York: Springer–Verlag.

Diez, D. M., Schoenberg, F. P. and Woody, C. D. (2010). *Analysis of neuronal responses to stimuli in cats using point process prototypes.* In review.

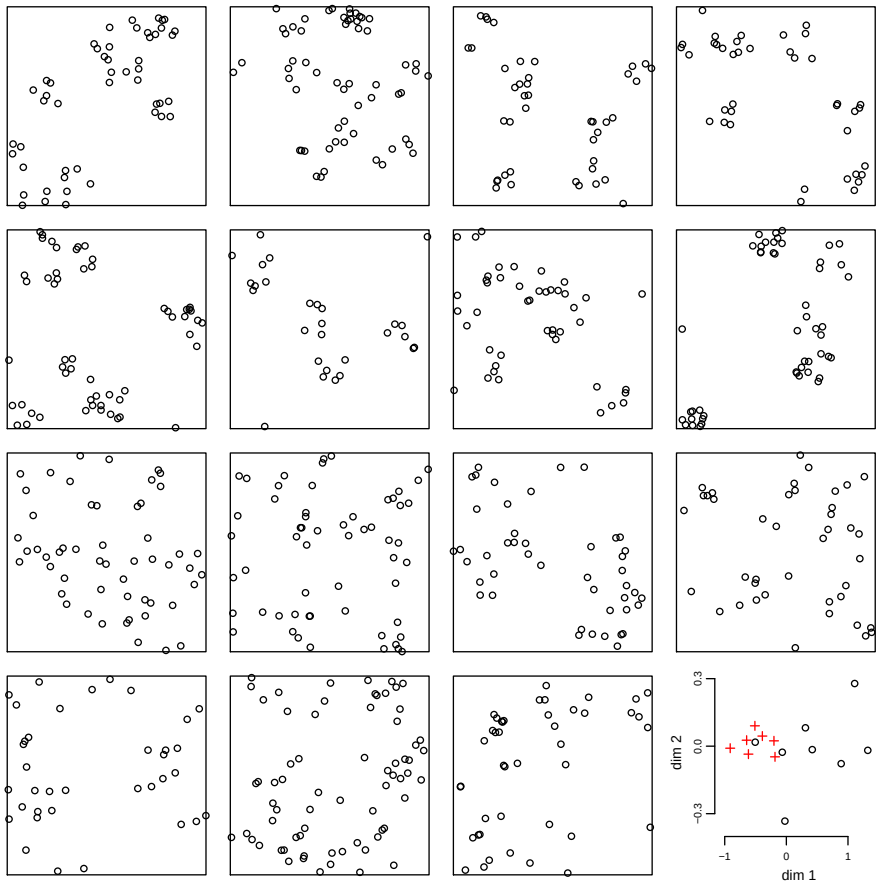

Fig. 7 The first two rows are realizations of a Neyman-Scott process with parent intensity of 8 and 6 children per parent uniformly distributed within a distance of 0.1. The point patterns in rows 3 and 4 are realizations of a Neyman-Scott process with parent intensity 24 and 2 children per parent within a radius of 0.1. The bottom-right panel shows MDS applied to these 15 patterns using the distance function D_K^2, where the black circles denote the eight patterns from the Neyman-Scott process with parameters (8,6,0.1) and the red crosses mark those of the Neyman-Scott process with parameters (24,2,0.1).

Fleming, T. R. and Harrington, D. P. (1981). A class of hypothesis tests for one and two sample censored survival data. *Communications in Statistics. Theory and Methods*, 10:763–794.

Lee, E. T. (1992). *Statistical Methods for Survival Data Analysis*. New York: Wiley.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of National Cancer Institute*, 22:719–748.

Schoenberg, F. P. and Tranbarger, K. E. (2008). Description of earthquake aftershock sequences using prototype point processes. *Environmetrics*, 19:271–286.

Tranbarger, K. E. (2005). *Point Process Prototypes, and Other Applications of Point Pattern Distance Metrics*. Ph.D. thesis, University of California, Los Angeles.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer.

Victor, J. and Purpura, K. (1997). Metric-space analysis of spike trains: theory, algorithms and application. *Journal of Neuroscience Methods*, 8:127–164.