Kernel Mean Embedding Based Hypothesis Tests for Comparing Spatial Point Patterns

Raif M. Rustamov and James T. Klosowski, AT&T Labs Research

Abstract

This paper introduces an approach for detecting differences in the first-order structures of spatial point patterns. The proposed approach leverages the kernel mean embedding in a novel way by introducing its approximate version tailored to spatial point processes. While the original embedding is infinite-dimensional and implicit, our approximate embedding is finite-dimensional and comes with explicit closed-form formulas. With its help we reduce the pattern comparison problem to the comparison of means in the Euclidean space. Hypothesis testing is based on conducting t-tests on each dimension of the embedding and combining the resulting p-values using the harmonic mean p-value combination technique. The main advantages of the proposed approach are that it can be applied to both single and replicated pattern comparisons, and that neither bootstrap nor permutation procedures are needed to obtain or calibrate the p-values. Our experiments show that the resulting tests are powerful and the p-values are well-calibrated; two applications to real world data are presented.

1 Introduction

Comparison of spatial point patterns is of practical importance in a number of scientific fields including ecology, epidemiology, and criminology. For example, such comparisons may reveal differential effects of the environment on plant species spread, uncover spatial variation in disease risk, or detect seasonal differences in crime locations. While exploratory analyses are vital for obtaining deep insights about pattern differences, such analyses can be subjective unless supplemented with formal hypothesis tests.

In this paper we are interested in comparing the first-order structures of point patterns. To be more precise, consider two point processes \( P \) and \( Q \) over the region \( A \subset \mathbb{R}^2 \) with the first-order intensities given by \( \lambda^P(\cdot) \) and \( \lambda^Q(\cdot) \). Given realizations from these processes, we would like to detect whether there are statistically significant differences in the first-order intensities. However, testing for equality, \( \lambda^P(\cdot) = \lambda^Q(\cdot) \), is not flexible enough. For example, when studying the spatial variation in disease risk, the diseased population is only a small fraction compared to the control population; naturally, the corresponding observed patterns will differ significantly in the overall counts of points—yet this is irrelevant to the substantive question. The more appropriate null hypothesis posits that there exists a constant \( c \) such that \( \lambda^P(\cdot) = c\lambda^Q(\cdot) \). Equality within a constant factor means that the intensities have the same functional form of spatial variation. To avoid dealing with the nuisance parameter \( c \), one can normalize the intensities to integrate to 1, giving rise to probability distributions \( p(\cdot) \) and \( q(\cdot) \) over the region \( A \); in [16] these are called the densities of event locations. Now, our null hypothesis is equivalent to the equality \( p(\cdot) = q(\cdot) \), which is an instance of the two-sample hypothesis testing problem (see, e.g. [3]).

In practice it is desirable to have nonparametric hypothesis testing approaches to pattern comparison that: a) capture a particular aspect of difference; b) can be applied to both single
and replicated patterns; and c) do not depend on resampling methods for (re-)calibration. Early nonparametric tests for pattern comparison \cite{12, 21} probe for differences in the $K$-functions \cite{33} of point patterns. Being based on a second-order property, the detected differences conflate the spatial variation in intensities with the interaction properties. Concentrating on the first-order properties, \cite{28, 27, 10} estimate the logarithm of the ratio between the intensities using kernel density estimation. Other approaches rely on counting events within pre-specified areas \cite{4, 2}. The recent work \cite{16} detects differences in the first-order structure by looking at the $L^2$-distance between kernel density estimates of the probability distributions above (i.e. $p(\cdot)$ and $q(\cdot)$). All of these first-order comparison approaches are limited to single patterns, and in practical situations most of them are calibrated with resampling methods.

In this paper, we introduce an approach that leverages the kernel mean embedding (KME) \cite{18, 30} to test for the equality $p(\cdot) = q(\cdot)$, which allows us to detect differences in the first-order structure of point patterns. Our approach is based on introducing an approximate version of the kernel mean embedding, aKME. While the original KME is infinite-dimensional and implicit, our approximate kernel mean embedding is finite-dimensional and comes with explicit closed-form formulas. With the help of aKME, we reduce the pattern comparison problem to the comparison of means in the Euclidean space. We carry out the latter by applying $t$-tests on each coordinate of the embedding, and obtain an overall $p$-value using the harmonic mean $p$-value combination technique \cite{17, 37}, leading to well-calibrated and powerful tests as confirmed by the simulations. The main advantages of the proposed approach are that it can be applied to both single and replicated pattern comparisons, and that neither bootstrap nor permutation procedures are needed to obtain or calibrate the $p$-values.

The ideas developed in this paper are in line with the recent surge of interest in applying reproducing kernel Hilbert space techniques to the comparison of probability distributions. For example, the Maximum Mean Discrepancy (MMD) is a measure of divergence between distributions \cite{18} which has already found numerous applications in statistics and machine learning. This has led to the general approach of kernel mean embeddings, see for example the recent review \cite{30} and citations therein. Our approximate embedding has its roots in the Random Fourier Features \cite{32}, and its application to the MMD \cite{41}, although the scheme proposed here is tailored to the two-dimensional setting, and has the ability to provide high-order approximations. There has already been some interest in applying the reproducing kernel methodology to spatial point processes, the roots going back to the 1980s \cite{7, 34} and more recently in \cite{15, 26, 39}. We discuss some of the connections between reproducing kernel machinery and kernel density estimation based methods commonly used with spatial point patterns in Section \ref{section:preliminaries}.

The main contributions of this paper are the proposed approximate kernel mean embedding (Section \ref{section:methodology}) and the hypothesis testing framework for comparison of point patterns (Section \ref{section:testing}). After investigating the empirical properties of the resulting tests on simulated data (Section \ref{section:simulations}), we present applications of the methodology to two real world datasets (Section \ref{section:applications}).

\section{Preliminaries}

\textbf{Kernel Mean Embedding} Mathematically rigorous development of the kernel mean embedding requires the machinery of the reproducing kernel Hilbert spaces, and the interested reader is referred to \cite{30}. For our purposes, it will be sufficient to concentrate on the dual version of the definition where the kernel mean embedding is expressed in terms of the feature maps.

Given a data instance $\mathbf{x} \in \mathcal{X}$ (in our context $\mathbf{x}$ will be a point in some region of $\mathbb{R}^2$), a nonlinear transformation $\phi : \mathcal{X} \to \mathcal{F}$ can be used to lift this point into a feature space $\mathcal{F}$ that is usually...
high/infinite dimensional. When the entire dataset is lifted like this, the hope is that a good feature map would reveal structures in the dataset that were not apparent in the original space $\mathcal{X}$. For example, in the context of classification in machine learning, two classes that are hard to separate in the original space may turn out linearly separable in the feature space. This is the foundation of kernel based approaches in machine learning (see e.g. [23]), with the idea going back to [9].

Assuming that an inner product $\langle \cdot, \cdot \rangle$ is available on $\mathcal{F}$, with each feature map $\phi$ one can associate the kernel function $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ defined by $k(x, y) = \langle \phi(x), \phi(y) \rangle$. It turns out that the converse is true as well: for each positive semi-definite (psd) kernel function $k(\cdot, \cdot)$ there exist a corresponding inner-product space $\mathcal{F}$ together with a feature map $\phi$ satisfying $k(x, y) = \langle \phi(x), \phi(y) \rangle$.

For a given psd kernel $k(\cdot, \cdot)$, the kernel mean embedding (KME) of a probability distribution $p$ over $\mathcal{X}$ is defined as the expectation $\mu_p = \mathbb{E}_{x \sim p}[\phi(x)] \in \mathcal{F}$.

An important property of this embedding is that when $k(\cdot, \cdot)$ is a characteristic kernel, the embedding is injective [30]. In other words, for two distributions $p$ and $q$ over $\mathcal{X}$, the equality $\mu_p = \mu_q$ is satisfied if and only if $p = q$. This means that in the context of two-sample hypothesis testing, intuitively, the problem of testing whether $p = q$ can be reduced to testing $\mu_p = \mu_q$, albeit in the infinite-dimensional space $\mathcal{F}$.

**Two-Sample Tests with KME** To avoid the infinite-dimensionality and absent an explicit formula for $\phi$, one can use the kernel trick to devise an actual hypothesis testing procedure. Consider the distance (as induced by the inner-product in $\mathcal{F}$) between KMEs of $p$ and $q$, which is known as the Maximum Mean Discrepancy (MMD) [18]:

$$\text{MMD}^2(p, q) = \|\mu_p - \mu_q\|^2.$$

The kernel trick, namely the application of the equality $k(x, y) = \langle \phi(x), \phi(y) \rangle$, allows to express the MMD in terms of $k(\cdot, \cdot)$ without requiring an explicit formula for the feature map $\phi$, namely,

$$\text{MMD}^2(p, q) = \mathbb{E}_{x, x' \sim p}[k(x, x')] - 2\mathbb{E}_{x \sim p, y \sim q}[k(x, y)] + \mathbb{E}_{y, y' \sim q}[k(y, y')]$$.

In practice this statistic is estimated from the samples by replacing the expectations with the sample means, with a nuance that the first and third terms can either include the diagonal terms or exclude them; the former gives a biased and the latter gives an unbiased estimator. Large enough values of $\text{MMD}^2(p, q)$ provide evidence against the null hypothesis of $p = q$.

This has been an extremely fruitful approach to two-sample testing as witnessed by the vast amount of the follow-up work, see the references in [18, 30]. The distribution of MMD under the null hypothesis is not asymptotically normal nor has a practically useful explicit form except in big data settings where one can compute an incomplete estimator of MMD that is asymptotically normal; these cases are described in generality in [38]. As a result, in typical scenarios some kind of approximation or resampling procedure is required to perform the test, see [18] and pointers therein.

**Connections to Kernel Density Estimation** The above described machinery has a close connection to kernel density estimation commonly used with spatial point patterns. For example,
let us consider the two-sample test statistic proposed in [3], which was used by [13] in a three-
dimensional setting and later adapted to the spatial point pattern comparison problem in [16].
This statistic is constructed by first obtaining kernel density estimates of the two distributions and
then computing the squared $L^2$-distance between the density estimates. Its connection to MMD
is described in [18, Section 3.3.1] and can be exemplified as follows: if, say, the kernel density
estimate is based on the Gaussian RBF kernel of width $\sigma$, then this statistic is equivalent to the
MMD computed with the same kernel but of width $\sqrt{2}\sigma$. Overall, MMD statistic is strictly more
general than the statistic of [3] because it allows using a larger variety of kernels; another benefit
is the stronger consistency guarantees that stem from the kernel mean embedding interpretation.
Similarly to the MMD, the statistic of [3] is not asymptotically normal under the null; nevertheless,
a normal approximation was suggested in [13]. However, this approximation leads to calibration
problems: [16] reported conservativeness of the resulting test, which led them to rely on bootstrap
for calibration.

Of course, instead of the $L^2$-distance, other distances can be computed between the kernel
density estimates to obtain statistics that capture complementary aspects of the differences be-
tween distributions. Similarly, one can realize that the MMD captures only one aspect of the
discrepancy between the kernel mean embeddings. For example, the kernel Fisher discriminant
analysis (KFDA) test statistic [14] is another statistic that can be computed via the kernel trick
and used to test the equality $\mu_p = \mu_q$. Similarly to the relationship between MMD and the
squared $L^2$-distance, the KFDA is connected to the $\chi^2$-distance between kernel density estimates.
An insightful review of this type of connections can be found in [22].

3 Approximate Kernel Mean Embedding

Instead of relying on the kernel trick, in this section we take an orthogonal path to avoiding the
infinite-dimensionality of the kernel mean embedding. Namely, specializing to the case $\mathcal{X} = \mathbb{R}^2$,
we construct a finite-dimensional approximate feature map $\phi : \mathbb{R}^2 \to \mathbb{R}^D$ such that $k(x, y) \approx
\phi(x)^\top \phi(y)$. As a result, testing $p = q$ can be reduced to testing $\mu_p = \mu_q$ in the
$D$-dimensional Euclidean space.

For the rest of the paper we will concentrate on the Gaussian RBF kernel $k(x, y) = e^{-\|x-y\|^2/2}$,
where $x, y \in \mathbb{R}^2$. To avoid clutter in the derivations the kernel width $\sigma$ is assumed to be 1; the
case of general $\sigma$ is mentioned below. Our goal is to construct an approximate feature map for
the Gaussian RBF kernel tailored to the two-dimensional case. As a starting point, we will use the
following representation for the Gaussian kernel (see, e.g. [31]),

$$
k(x, y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{w^\top w}{2} \eta(w^\top x)^\top \eta(w^\top y)} dw,
$$

where $\eta(\cdot) = [\cos(\cdot) \; \sin(\cdot)]^\top$. To compress the notation, let $f(w) = \eta(w^\top x)^\top \eta(w^\top y)$ where the dependence on $x$ and $y$ is suppressed from the notation. We change to the radial coordinates, and notice that $f$ is an even function to obtain:

$$
k(x, y) = \frac{1}{2\pi} \int_{0}^{2\pi} \int_{0}^{\infty} e^{-\frac{r^2}{2} f(rz_\theta)} dr d\theta = \frac{1}{\pi} \int_{0}^{\pi} \int_{0}^{\infty} e^{-\frac{r^2}{2} f(rz_\theta)} dr d\theta,
$$

where $z_\theta$ is a unit vector in $\mathbb{R}^2$ making an angle of $\theta$ with the $x$-axis. We approximate the integration with respect to $\theta$ by the simple mean over a set of equidistant $\theta_i = (i - 1)\pi/m$,
Figure 1: Approximate embedding can be seen as projecting the point pattern onto a line followed by the application of sin/cos functions with the appropriate frequencies. The overall dimensionality of the resulting embedding is \( D = 2 \times \text{number of projections} \times \text{order of quadrature} \). Figure adapted from [32].

For a given integer \( \ell \geq 1 \), denoting the quadrature roots by \( \{r_j\}_{j=1}^\ell \) and the corresponding weights by \( \{\omega_j\}_{j=1}^\ell \), we obtain

\[
  k(x, y) \approx \sum_{i=1}^m \sum_{j=1}^\ell \omega_j f(r_j v_i) = \sum_{i=1}^m \sum_{j=1}^\ell \omega_j \eta(r_j v_i^\top x) \eta(r_j v_i^\top y).
\]

The crucial point is that the terms containing \( x \) and \( y \) are untangled, which allows us to write \( k(x, y) \) approximately as a dot product in the embedding space: \( k(x, y) \approx \phi(x)^\top \phi(y) \), where \( \phi : \mathbb{R}^2 \to \mathbb{R}^D, \ D = 2m\ell \) is explicitly given by

\[
  \phi(x) = [\ldots, \sqrt{\omega_j} \sin(r_j v_i^\top x), \sqrt{\omega_j} \cos(r_j v_i^\top x), \ldots]^\top.
\]

Figure 1 (adapted from [32]) gives an intuitive depiction of this construction. An R implementation of the embedding step can be found in Appendix A. The circular ensemble interpretation of [41] applies to our construction as well, giving further insights and clarifying the relationship to the MMD.

While there is a clear connection to the Random Fourier Features [32], yet our construction is tailored to the two-dimensional setting and is based on a deterministic quadrature rule that quickly provides a high-order approximation. In contrast, the construction in [32] and its improvements [5, 40, 31] are geared towards high-dimensional data and, thus, depend on stochasticity to avoid the curse of dimensionality. Nevertheless, the hypothesis tests proposed in Section 4 do generalize to such embeddings and, consequently, can be applied in high-dimensional settings.

---

\[1\] The values of the roots and weights can be retrieved from the following URL: http://www.jaeckel.org/RootsAndWeightsForRadialGaussHermiteQuadrature.cpp
We note that the case of the Gaussian kernel with the width $\sigma$ can be accommodated by replacing $x$ with $x/\sigma$. To avoid committing to a single $\sigma$, in practice, we construct multiple feature maps corresponding to different values of the kernel width, and concatenate them together.

Finally, having introduced the approximate feature map, we can define the approximate kernel mean embedding (aKME) of a probability distribution $p(\cdot)$ over the region $A \subset \mathbb{R}^2$ as $\mu^p = \mathbb{E}_{x \sim p}[\phi(x)] \in \mathbb{R}^D$. From a theoretical point of view, this approximate construction loses the injectivity of the kernel mean embedding, and, consequently, the consistency of the test. However, as we will see in the experimental section, this does not lead to any apparent issues with the testing power in practical situations. We believe that this is related to the approximation quality of the proposed embedding whereby one is able to capture most of the information about the distribution in a small number of dimensions.

3.1 Spatial Point Pattern aKME

Let $P$ be a point process on the region $A \subset \mathbb{R}^2$ with the first-order intensity function given by $\lambda^P(\cdot)$. The intensity function of $P$ can be normalized to give the probability distribution $p(\cdot)$ on $A$:

$$p(x) = \frac{1}{\int_A \lambda^P(x) \, dx} \lambda^P(x), \quad x \in A. \quad (3.1)$$

In [16] this is called the density of event locations. The aKME of the point process $P$ is defined via the embedding of $p(\cdot)$, namely, $\mu^P = \mu^p = \mathbb{E}_{x \sim p}[\phi(x)] \in \mathbb{R}^D$. In practice, we only have access to a point process via its realizations, and so, we must be able to estimate the aKME in an unbiased manner from a point pattern; this is the focus of the current subsection.

Consider a realization $X = \{x_1, x_2, \ldots, x_n\}$ of the point process $P$. The estimator that we will be interested in is the one that replaces the expectation appearing in the aKME formula by the sample mean:

$$\hat{\mu}^X = \frac{1}{n} \sum_{j=1}^{n} \phi(x_j) \in \mathbb{R}^D. \quad (3.2)$$

We require this estimate to be unbiased, $\mathbb{E}_{X \sim P}[\hat{\mu}^X] = \mu^P$, where the expectation is taken over the realizations of the process $P$. Unbiasedness is essential as it ensures that the estimate $\hat{\mu}^X$ captures information about the functional form of the intensity function rather than some other properties of the point process.

For Poisson processes it is easy to see that unbiasedness holds, yet writing out an explicit proof will make clear what properties of the point process are relevant to unbiasedness. To this end, let $P$ be a Poisson process with intensity $\lambda^P(x) = \beta p(x)$, where the constant $\beta$ is the integral in the denominator of Eq. (3.1). We will condition on the number of the points in the realization, $n$. As well-known, this can be seen as a Binomial process on $A$, which we denote by $B_n$; the intensity of $B_n$ is given by $np(x)$. We can write,

$$\mathbb{E}_{X \sim P}[\hat{\mu}^X] = \mathbb{E}_{n \sim \text{Pois}(\beta), X \sim B_n} \left[ \frac{1}{n} \sum_{j=1}^{n} \phi(x_j) \right] = \mathbb{E}_{n \sim \text{Pois}(\beta)} \left[ \frac{1}{n} \mathbb{E}_{X \sim B_n} \left[ \sum_{j=1}^{n} \phi(x_j) \right] \right].$$

Using Campbell’s formula we have

$$\frac{1}{n} \mathbb{E}_{X \sim B_n} \left[ \sum_{j=1}^{n} \phi(x_j) \right] = \frac{1}{n} \int_A \phi(x) np(x) \, dx = \int_A \phi(x) p(x) \, dx.$$
Substituting this back, we get
\[
E_{\mathbf{x} \sim P}[\hat{\mu}^X] = E_{n \sim \text{Pois}(\beta)} \left[ \int_A \phi(x)p(x)dx \right] = \int_A \phi(x)p(x)dx = E_{\mathbf{x} \sim P}[\phi(x)] = \mu^p,
\]
which proves unbiasedness.

The crucial ingredient in the derivation is the following: conditioning on the number of points in the pattern does not have an effect on the functional form of the intensity. For example, above it went from \( \beta p(x) \) to \( np(x) \), which is a change by a constant factor. While theoretically it is possible to construct point processes where this does not hold, the simplicity of this requirement makes it possible to use domain knowledge to assess how likely is the condition to hold in practice. For example, in plant biology applications the functional form of the intensity is determined by the environmental factors such as light conditions, level of water underground, and quality of soil, which makes it plausible to assume that the condition holds.

The non-Poisson case is relevant to the proposed hypothesis test for replicated pattern comparison. Different replications commonly have differing numbers of points. If the condition above is not satisfied—the point pattern properties depend profoundly on the number of points—then the replicated pattern comparison cannot be performed by any test, unless all of the involved patterns have exactly the same number of points. In this latter case our test can be used too because for a fixed number of points unbiasedness holds directly by Campbell’s formula.

### 4 Comparing Spatial Point Patterns with aKME

Consider two point processes \( P \) and \( Q \) in \( A \subset \mathbb{R}^2 \) with the first-order intensity functions given by \( \lambda^P(\cdot) \) and \( \lambda^Q(\cdot) \). We would like to test the null hypothesis of whether there exists a constant \( c \) such that \( \lambda^P(\cdot) = c\lambda^Q(\cdot) \). Equality within a constant factor means that the intensities of the two processes have the same functional form. This is different from testing \( \lambda^P(\cdot) = \lambda^Q(\cdot) \) because our null hypothesis can hold true even if the realizations from \( P \) and \( Q \) have vastly differing numbers of points. We can compute the probability distributions \( p(\cdot) \) and \( q(\cdot) \) from the intensities \( \lambda^P(\cdot) \) and \( \lambda^Q(\cdot) \) using the formula Eq. (3.1). Clearly, the null hypothesis above can be written as \( p = q \), allowing us to use the aKME machinery and reducing the problem to testing whether \( \mu^p = \mu^q \) holds in the \( D \)-dimensional Euclidean space.

**Single Pattern Comparison** Assume that we observe two point patterns \( X = \{x_1, x_2, \ldots, x_{n_1}\} \) and \( Y = \{y_1, y_2, \ldots, y_{n_2}\} \) from two Poisson processes \( P \) and \( Q \) on the region \( A \subset \mathbb{R}^2 \) with the true intensity functions given by \( \lambda^P(\cdot) \) and \( \lambda^Q(\cdot) \). We would like to test the null hypothesis of whether there exists a constant \( c \) such that \( \lambda^P(\cdot) = c\lambda^Q(\cdot) \).

Applying the aKME machinery, we reduce the test to checking whether \( \mu^p = \mu^q \) in the \( D \)-dimensional Euclidean space. While the classic Hotelling’s \( T^2 \) test seems like a natural choice, yet in practice even the computation of the precision matrix turns out to be unstable due to the nonlinear relationships between the coordinates of the approximate feature map. This is why we perform the test by applying \( t \)-tests independently to each coordinate and combining the resulting \( p \)-values as explained below.

More precisely, for each coordinate \( i = 1, 2, \ldots, D \), by Poisson assumption we can consider \( \{\phi_i(x_j), x_j \in X\} \) and \( \{\phi_i(y_j), y_j \in Y\} \) to be two independent samples of sizes \( n_1 \) and \( n_2 \) respectively; here \( \phi_i \) is the \( i \)-th coordinate of the approximate feature map. Our goal is to compare the sample means \( \hat{\mu}_i^X \) and \( \hat{\mu}_i^Y \). Note that since the approximate feature maps are expressed in terms of sines and cosines, their range is bounded, and so the classical Central Limit Theorem
can be applied to deduce the approximate normality of the sample means $\hat{\mu}_i^X$ and $\hat{\mu}_i^Y$ when the sample sizes are big enough. As a result, we can apply the Behrens-Fisher-Welch $t$-test (without assuming equality of variances) to obtain the corresponding $p$-value $p_i$.

All together we end up with a set of $p$-values $p_1, p_2, ..., p_D$ one per coordinate of the aKME. In principle, one can resort to multiple testing procedures to reject or retain the global null. However, it is useful to compute an overall $p$-value for the the global null hypothesis $\mu^P = \mu^Q$ so that it can be used in downstream multiple testing procedures. To obtain such an overall $p$-value we use the recently popularized harmonic mean $p$-value combination approach \cite{17, 37}. Namely, one defines

$$\hat{p} = H \left( \frac{D}{\frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_D}} \right),$$

where $H(\cdot)$ is a function whose precise form is described in \cite{37}. While at the theoretical level this approach assumes independence between $p$-values, yet as shown in \cite{37} it is very robust when there are dependencies. We observed this to be true with our $p$-values as well: despite the dependencies, we found that $\hat{p}$ obtained in this manner resulted in well-calibrated tests with better power than simple alternatives such as the Bonferroni adjustment for multiple testing. Other $p$-value combination approaches such as Fisher’s and Stouffer’s methods (see, e.g. \cite{29, 24}) failed to give well-calibrated $p$-values likely due to their strong reliance on the independence assumption.

**Replicated Pattern Comparison** While the above test was based on the Poisson assumption, we consider another setting where the only assumption on the point processes is that the sample mean gives an unbiased estimate of aKME. Here, for two such general processes $P$ and $Q$ on the region $A \subset \mathbb{R}^2$ with true intensity functions given by $\lambda^P(\cdot)$ and $\lambda^Q(\cdot)$, we observe a set of patterns $P^* = \{X_1, X_2, ..., X_{m_1}\}$ and $Q^* = \{Y_1, Y_2, ..., Y_{m_2}\}$ independently drawn from $P$ and $Q$ respectively. Our goal is to test whether $\lambda^P(\cdot) = c\lambda^Q(\cdot)$, which is reduced to testing $p = q$ exactly as before.

In this setting, we can consider $\{\hat{\mu}_i^{X_k}, X_k \in P^*\}$ and $\{\hat{\mu}_i^{Y_k}, Y_k \in Q^*\}$ as independent samples of sizes $m_1$ and $m_2$. This allows us to test for the equality of $i$-th coordinates $\mu_i^P = \mu_i^Q$ using the $t$-test; here, the validity of the $t$-test stems from independence of replications and holds for general processes; this should be contrasted to the above test which derives its validity from the independence of samples in the Poisson process. The set of $D$ $p$-values obtained from the $t$-tests are combined once again using Eq. (4.1) to obtain an overall $p$-value $\hat{p}$.

We finish with an important remark about the nature of this test. The point processes $P$ and $Q$ do not have to be of the same type for the null hypothesis to hold. The test is concerned only with the functional form of the intensity, and not the type of the process. For example, if $P$ is a homogeneous inhibition process and $Q$ is a homogeneous cluster process, then the null hypothesis still holds due to the first-order intensity functions being the same or differing by a constant factor.

5 Experiments

Our goal in this section is to investigate the size and the power of the proposed tests. We also demonstrate two applications to real world data. The aKME embedding is constructed using four radial projections and four roots for the polar Gauss-Hermite formula (i.e. $m = 4$, $\ell = 4$, and $D = 32$ in the notation of Section 3) resulting in $D = 32$. To avoid the selection of the kernel width parameter, we concatenate together aKMEs corresponding to $\sigma = 1/16, 1/8,$ and $1/4$, when the point pattern domain is the unit square; the dimensionality of the concatenated aKME is $96 = 3 \times 32$. 

8
| Model | Mean Point Count | $\gamma$ | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.1$ | Guan, $\alpha = 0.1$ |
|-------|------------------|---------|-----------------|-----------------|-----------------|---------------------|
| Size  | CSR              | 100     | 0.012           | 0.060           | 0.105           |                     |
|       |                  | 400     | 0.012           | 0.056           | 0.099           |                     |
|       | Linear           | 100     | 1               | 0.010           | 0.054           | 0.096              |
|       |                  |         | 2               | 0.014           | 0.060           | 0.110              |
|       |                  | 400     | 1               | 0.008           | 0.046           | 0.088              |
|       |                  |         | 2               | 0.008           | 0.050           | 0.090              |
|       | Sine             | 100     | 1               | 0.012           | 0.056           | 0.105              |
|       |                  |         | 2               | 0.012           | 0.047           | 0.085              |
|       |                  | 400     | 1               | 0.010           | 0.054           | 0.100              |
|       |                  |         | 2               | 0.007           | 0.053           | 0.095              |
| Power | CSR              | 100     | 0.644           | 0.746           | 0.790           |                     |
|       |                  | 400     | 0.924           | 0.966           | 0.978           |                     |
|       | Linear           | 100     | 1               | 0.550           | 0.627           | 0.678              |
|       |                  |         | 2               | 0.690           | 0.754           | 0.806              |
|       |                  | 400     | 1               | 0.844           | 0.918           | 0.944              |
|       |                  |         | 2               | 0.912           | 0.956           | 0.971              |
|       | Sine             | 100     | 1               | 0.780           | 0.873           | 0.909              |
|       |                  |         | 2               | 0.815           | 0.886           | 0.916              |
|       |                  | 400     | 1               | 1.000           | 1.000           | 1.000              |
|       |                  |         | 2               | 1.000           | 1.000           | 1.000              |

Table 1: Rejection rates for the single pattern comparison test when the Poisson assumption holds.

### 5.1 Simulations

**Single Pattern Comparison**  Our simulation setting for single pattern comparison closely follows the design of [20]. We generate a homogeneous Poisson process (CSR) and inhomogeneous Poisson processes (Linear and Sine) on the unit square $[0, 1] \times [0, 1]$. For the inhomogeneous processes the intensity functions are given by $\exp(-\gamma x)$ and $\exp(-\gamma \sin(2\pi x))$, where $x$ is the abscissa and $\gamma$ takes the values of $\gamma = 1, 2$. All of the intensity functions are normalized so that the expected number of points per realization is roughly 100 and 400. Each type of model is simulated 2000 times.

Table 1 lists rejection rates at the nominal levels $\alpha = 0.01, 0.05, 0.1$. The top part of the table corresponds to the case where the patterns being compared come from the same intensity model. The resulting rejection rates give the size of the test; note that all of them are close to the nominal sizes, which confirms that our test is well-calibrated. In the bottom half of the table the patterns being compared come from different intensity models; the corresponding rejection rates give the power of the test. For comparison, in the last column we include the rejection rates at $\alpha = 0.1$ level taken from [20] (specifically, Table 1 with the choice of $t = 0.2$; the values for CSR were not available). Note that while our test compares two point patterns, Guan’s test is a goodness-of-fit test, whereby it compares a pattern against a model, and so it has access to the theoretical quantities, which, in principle, gives it a competitive advantage. Nevertheless, it can be seen from the table that our test compares very favorably to Guan’s test in terms of power.

Next we investigate what happens when the Poisson assumption is violated. To this end we run comparisons between patterns from homogeneous Poisson process and patterns from homogeneous...
Table 2: Single pattern testing depends on the validity of the Poisson assumption. When it is violated, the test size does not match the nominal rate; inhibition and clustering have opposite effects on the size.

| Model     | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.1$ |
|-----------|-----------------|-----------------|-----------------|
| **Size**  |                 |                 |                 |
| Hardcore-1| 0.012           | 0.046           | 0.090           |
| Hardcore-2| 0.008           | 0.035           | 0.073           |
| Hardcore-3| 0.002           | 0.009           | 0.020           |
| Cluster-1 | 0.056           | 0.164           | 0.271           |
| Cluster-2 | 0.160           | 0.372           | 0.508           |
| Cluster-3 | 0.413           | 0.686           | 0.780           |

Table 3: Size of the single pattern comparison test when the Poisson assumption is violated due to clustering. Using the effective sample size in the $t$-tests for clustering processes results in sizes that are close to the nominal rate.

| Model     | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.1$ |
|-----------|-----------------|-----------------|-----------------|
| **Size**  |                 |                 |                 |
| Cluster-1 | 0.011           | 0.051           | 0.086           |
| Cluster-2 | 0.014           | 0.048           | 0.086           |
| Cluster-3 | 0.015           | 0.050           | 0.078           |

non-Poisson processes. This is the null case, because the functional form of intensity is the same (constant) for all of these cases. The non-Poisson processes we consider are generated using the spatstat package[6] and are as follows. We use Matern’s Model II inhibition process with the inhibition distance parameter setting of $r = 0.01, 0.02, 0.04$; these are referred to as Hardcore-1,2,3. We also generate realizations from Matern’s cluster process with the radius parameter value of 0.1, and with the mean number of points parameter $\mu = 1, 2, 4$; these are called Cluster-1,2,3 respectively. The underlying intensity for all of the processes is set to yield an average of 100 points per pattern.

Table 2 lists the resulting null rejection rates which correspond to the size of the test. We see that the inhibition process results in the test becoming more conservative; we conjecture that this is true in general, and so, when inhibition can be argued based on the domain knowledge, the rejections obtained via our test are still valid. In contrast, clustering quickly leads to an anti-conservative test. The latter is not surprising, as the clustered points cannot be seen as independently drawn. One way of looking at this is that the effective sample size[19, 35, 1] in a clustered process is lower than the number of points in the pattern. Next, we investigate whether correcting the $t$-tests for the sample size can bring the test size to the nominal level. We made a back-of-the-envelope computation of the effective sample size by the formula: $n(1 - \exp(-\mu))/\mu$; note that this is different from the naive $n/\mu$ due to clusters with zero points being deleted. The resulting sizes are shown in Table 3 and match the nominal rates more closely. Of course, in practice, one pattern is not enough to deduce whether the clustering is due to the heterogeneity of the intensity or due to clustering; thus, some modeling assumptions on the point process would be required in order to carry out such a correction.
\( \alpha = 0.01 \quad \alpha = 0.05 \quad \alpha = 0.1 \)

| Size  | 0.008 | 0.045 | 0.090 |
|-------|-------|-------|-------|
| Power | 0.764 | 0.900 | 0.938 |

Table 4: Rejection rates for the replicated pattern comparison test.

**Replicated Pattern Comparison**  Since in the replicated pattern test does not require the Poisson assumption, we generate seven classes of point patterns first homogeneous and second non-homogeneous, yielding fourteen types of patterns in total. For homogeneous versions, we generate CSR, Hardcore-1,2,3, and Cluster-1,2,3. To obtain the non-homogeneous versions of these, we start with a homogeneous version and independently thin it with spatially varying retention probability of \( \exp(-x) \); this is based on [36]. As a result, we obtain inhomogeneous point patterns where the intensity has the functional form \( \exp(-x) \). All generation processes were manipulated so that the expected number of points per realization is roughly 100. The first row in the Table 4 lists the rejection rates corresponding to the cases when the patterns compared have the same (in-)homogeneity type without regard for the class of the pattern. This corresponds to the null case, and we can see that the sizes are close to the nominal levels. The second row lists the rejection rates when the (in-)homogeneity types differ, and so correspond to the power of the test.

### 5.2 Real-World Data

**Single Pattern: Cancer Data**  We apply the single pattern comparison test to a dataset from an epidemiological study relating to the locations of larynx and lung cancer occurrences in Chorley-Ribble area of Lancashire, England during the years of 1974-1983; the source of the data is [11]. Figure 2 plots the 58 cases of larynx cancer and 978 cases of lung cancer together with the location of an industrial incinerator in this area. Following [11], we assume that the inhomogeneous Poisson process model applies, and that the distribution of the lung cancer cases can be used as a surrogate for the susceptible population. Let the corresponding true intensity functions be \( \lambda_{\text{Larynx}}(\cdot) \) and \( \lambda_{\text{Lung}}(\cdot) \). If there is an effect of the relative location of the incinerator on larynx cancer (and not on lung cancer), then the functional forms of these intensities would be different, i.e. there would be some non-constant function \( \rho(\cdot) \) such that \( \lambda_{\text{Larynx}}(\cdot) = \rho(\cdot) \lambda_{\text{Lung}}(\cdot) \). If, furthermore, \( \rho(\cdot) \) depends on the distance to the incinerator, this would be an evidence of a differential effect of the incinerator on the two types of cancer.

Considering the larynx cancer locations as the point pattern \( X \) and the lung cancer locations as the pattern \( Y \), we would like to test whether the functional form of the intensity is the same for these patterns. Note that the null hypothesis states that there is a constant \( c \) such that the intensities of the two point processes satisfy \( \lambda_{\text{Larynx}}(\cdot) = c \lambda_{\text{Lung}}(\cdot) \). Applying our methodology, we find the \( p \)-value of \( \hat{\rho} = 0.905 \), which leads us to retain the null hypothesis; this is in agreement with the re-analyses of this data such as [12].

**Replicated Patterns: Chicago Crime**  In this experiment we use the dataset provided by Chicago Data Portal\(^2\) that reflects reported incidents of crime that occurred in the City of Chicago. We consider all of the crimes during the year of 2018 within a geospatial window (shown in Figure 3), and the goal is to compare the spatial patterns of each crime type between weekdays and
Figure 2: Locations of larynx (red dots) and lung (blue pluses) cancers together with the location of the industrial incinerator (black circle cross).

Figure 3: Geospatial window used for the Chicago Crime experiment. The window coincides with the OpenStreetMap tile at zoom level of 10 that covers the city of Chicago and the surrounding region. Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org.
Table 5: Application of the replicated pattern comparison to Chicago Crime dataset. The entries in bold correspond to the rejected hypotheses with the BH procedure at the FDR level of 0.1.

| Crime Type       | Tuesday #Days | Thursday #Days | Saturday #Days | Tue vs Thu p-value | Tue vs Sat p-value |
|------------------|---------------|----------------|----------------|-------------------|-------------------|
| Theft            | 40            | 144.7          | 40             | 150.3             | 0.125             | **0.001**        |
| Criminal Damage  | 40            | 88.0           | 40             | 115.5             | 0.179             | **0.002**        |
| Deceptive Practice| 40            | 43.0           | 40             | 36.2              | 0.825             | **0.006**        |
| Battery          | 40            | 50.5           | 40             | 61.4              | 0.566             | **0.014**        |
| Narcotics        | 40            | 19.9           | 40             | 22.9              | 0.099             | **0.051**        |
| Other Offense    | 40            | 28.1           | 40             | 30.1              | 0.915             | 0.613            |
| Motor Veh. Theft | 40            | 25.9           | 40             | 22.8              | 0.475             | 0.657            |
| Assault          | 40            | 40.3           | 40             | 37.9              | 0.987             | 0.725            |
| Burglary         | 40            | 32.8           | 40             | 28.1              | 0.788             | 0.968            |

weekends. We wanted to pick days that are maximally spaced out and avoid ambiguity (e.g. Friday nights), so we decided to compare the following days of the week: Tuesday, Thursday, and Saturday. The comparison between Tuesday and Thursday serves as a sanity check, as we do not expect to see any differences between them.

For each day and type of crime we generate the crime patterns, and filter by the count of points keeping the patterns that have more than 20 points. Next we count how many distinct patterns are available for each day and type of crime; we kept the crime types that had at least 20 patterns for each of Tuesday, Thursday, and Saturday. The types of crimes remaining after this filtering are shown in Table 5 together with the number of days and the average count of points per pattern.

The results of comparisons are shown in the last two columns of Table 5. We run the Benjamini-Hochberg [8] procedure on the 18 resulting p-values at the false discovery rate of 0.1, and the rejected hypotheses are indicated by the p-values in bold. As expected, no differences were detected between Tuesday and Thursday patterns. On the other hand, we see that there are statistically significant differences between Tuesday and Saturday patterns in the following categories of crime: theft, criminal damage, deceptive practice, and battery.

6 Conclusion

We have introduced an approach to detect differences in the first-order structure of spatial point patterns. The proposed approach leverages the kernel mean embedding in a novel way, by introducing its approximate version. Hypothesis testing is based on conducting t-tests on each dimension of the approximate embedding and combining them using the harmonic mean approach. Our experiments confirm that the resulting tests are powerful and the p-values are well-calibrated. Two applications to real world data have been presented.

An interesting avenue for future research would be to leverage the directionality of the projections that correspond to the subsets of dimensions of the aKME (c.f. Figure 1). One can envision a technique where anisotropy in the first-order structure difference would be detected by appropriately combining the p-values. Another aspect would be to investigate the scales at which the differences happen, this time combining the p-values for each scale. The harmonic mean approach is suitable for such multi-facet and multi-level analyses as shown in [37]. Another direc-
tion is studying the proposed approach in the context of goodness-of-fit testing of the first-order structure. Since our approach allows for vastly differing numbers of points in the patterns, one can compute a low-variance approximation to the theoretical aKME of the model distribution by constructing a pattern via drawing a large number of points from that distribution. Finally, it would be desirable to extend the proposed technique to the high-dimensional setting, where the curse of the dimensionality would be an important factor that can affect the testing power.

References

[1] Jonathan Acosta, Ronny Vallejos, and Daniel Griffith. On the effective geographic sample size. *Journal of Statistical Computation and Simulation*, 88(10):1958–1975, 2018.

[2] M.V. Alba-Fernández, F.J. Ariza-López, M. Dolores Jiménez-Gamero, and J. Rodríguez-Avi. On the similarity analysis of spatial patterns. *Spatial Statistics*, 18:352 – 362, 2016.

[3] N.H. Anderson, P. Hall, and D.M. Titterington. Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *Journal of Multivariate Analysis*, 50(1):41 – 54, 1994.

[4] Martin A. Andresen. Testing for similarity in area-based spatial patterns: A nonparametric monte carlo approach. *Applied Geography*, 29(3):333 – 345, 2009.

[5] Haim Avron, Vikas Sindhwani, Jiyan Yang, and Michael W. Mahoney. Quasi-monte carlo feature maps for shift-invariant kernels. *Journal of Machine Learning Research*, 17(120):1–38, 2016.

[6] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London, 2015.

[7] Robert Bartoszynski, Barry W. Brown, Charles M. McBride, and James R. Thompson. Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary poisson process. *Ann. Statist.*, 9(5):1050–1060, 09 1981.

[8] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[9] T. M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, EC-14(3):326–334, June 1965.

[10] Tilman M. Davies and Martin L. Hazelton. Adaptive kernel estimation of spatial relative risk. *Statistics in Medicine*, 29(23):2423–2437, 2010.

[11] Peter J. Diggle. A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 153(3):349–362, 1990.

[12] Peter J. Diggle, Nicholas Lange, and Francine M. Benes. Analysis of variance for replicated spatial point patterns in clinical neuroanatomy. *Journal of the American Statistical Association*, 86(415):618–625, 1991.
[13] Tarn Duong, Bruno Goud, and Kristine Schauer. Closed-form density-based framework for automatic detection of cellular morphology changes. *Proceedings of the National Academy of Sciences*, 109(22):8382–8387, 2012.

[14] Moulines Eric, Francis R. Bach, and Zaïd Harchaoui. Testing for homogeneity with kernel fisher discriminant analysis. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 609–616. Curran Associates, Inc., 2008.

[15] Seth Flaxman, Yee Whye Teh, and Dino Sejdinovic. Poisson intensity estimation with reproducing kernels. *Electron. J. Statist.*, 11(2):5081–5104, 2017.

[16] I. Fuentes-Santos, W. González-Manteiga, and J. Mateu. A nonparametric test for the comparison of first-order structures of spatial point processes. *Spatial Statistics*, 22:240 – 260, 2017. Spatio-temporal Statistical Methods in Environmental and Biometrical Problems.

[17] I. J. Good. Significance tests in parallel and in series. *Journal of the American Statistical Association*, 53(284):799–813, 1958.

[18] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *J. Mach. Learn. Res.*, 13:723–773, March 2012.

[19] Daniel A. Griffith. Effective geographic sample size in the presence of spatial autocorrelation. *Annals of the Association of American Geographers*, 95(4):740–760, 2005.

[20] Yongtao Guan. A goodness-of-fit test for inhomogeneous spatial poisson processes. *Biometrika*, 95(4):831–845, 2008.

[21] Ute Hahn. A studentized permutation test for the comparison of spatial point patterns. *Journal of the American Statistical Association*, 107(498):754–764, 2012.

[22] Z. Harchaoui, F. Bach, O. Cappe, and E. Moulines. Kernel-based methods for hypothesis testing: A unified view. *IEEE Signal Processing Magazine*, 30(4):87–97, July 2013.

[23] Trevor J. Hastie, Robert John Tibshirani, and Jerome H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer series in statistics. Springer, New York, 2009.

[24] N A Heard and P Rubin-Delanchy. Choosing between methods of combining p-values. *Biometrika*, 105(1):239–246, 01 2018.

[25] Peter Jaeckel. A Note on Multivariate Gauss-Hermite Quadrature. http://www.btinternet.com/~pjaeckel/ANoteOnMultivariateGaussHermiteQuadrature.pdf, May 2005.

[26] Wittawat Jitkrittum, Wenkai Xu, Zoltan Szabo, Kenji Fukumizu, and Arthur Gretton. A linear-time kernel goodness-of-fit test. In *NeurIPS*, 2017.

[27] Julia E. Kelsall and Peter J. Diggle. Kernel estimation of relative risk. *Bernoulli*, 1(1/2):3–16, 1995.

[28] Julia E. Kelsall and Peter J. Diggle. Non-parametric estimation of spatial variation in relative risk. *Statistics in Medicine*, 14(21 - 22):2335–2342, 1995.
Here we provide an R implementation of the single pattern comparison test for the example demonstrated in Section 5.2, namely comparison of the larynx and lung cancer location patterns.
library(spatstat)
library(harmonicmeanp)

radialGaussHermiteData <- function(k) {
  if (k == 4) {
    Roots4 = c(
      0.3961205684809970482046989062,
      1.1767346714185921750900701015,
      2.201067629189581946543092946,
      3.4836109980286615716953100856
    )
    Weights4 = c(
      0.2279981086730596270303131898,
      0.538464759426994159094177725,
      0.2215779133653168055615931299,
      0.0119592185346294083139159552
    )
    list(x = Roots4, w = Weights4)
  } else {
    stop("Please refer to the values in http://www.jaeckel.org/RootsAndWeightsForRadialGaussHermiteQuadrature.cpp")
  }
}

get_aKME <- function(p,
sigmas,
  n_radial = 4,
  n_linear = 4) {
  # radial unit vectors
  theta = seq(0, pi, length.out = n_radial + 1)[-n_radial]
  W0 = rbind(cos(theta), sin(theta))

  # quadrature roots and weights
  temp = radialGaussHermiteData(n_linear)
  Roots = temp$x
  Weights = temp$w

  W = do.call('cbind', lapply(seq_along(Roots), function(i)
    Roots[i] * W0))
  coeff = rep(sqrt(Weights / n_radial), each = nrow(W0))

  # project the points on the lines, c.f. Figure 1
  tt = cbind(p$x, p$y) %*% W

  # embedding for each sigma
  emb = lapply(seq_along(sigmas), function(i)
    cbind(cos(tt / sigmas[i]) * coeff, sin(tt / sigmas[i]) * coeff))

  # concatenated embedding
  do.call('cbind', emb)
}

# load data
d = chorley
d1 = split(d)$lung
d2 = split(d)$larynx
We use several Gaussian widths at once:
\[
\text{sigma\_list0} = c(1/16, 1/8, 1/4) \quad \text{#these values are for } [0,1] \times [0,1] \text{ window}
\]
\[
\text{sigma\_list} = \text{diameter(\text{Window(d)})} \times \text{sigma\_list0} / \sqrt{2}
\]

\[
a\text{KME1} = \text{get\_aKME(d1, sigma\_list)}
\]
\[
a\text{KME2} = \text{get\_aKME(d2, sigma\_list)}
\]

#p-values for t-tests on each dimension of the embeddings
\[
p\text{-ttests} = \text{sapply(seq\_len(ncol(aKME1)), function(i)
   t.test(aKME1[, i], aKME2[, i])$p\_value)}
\]

#overall p-value
\[
p\_harmonic = p\_hmp(p\_ttests)
\]
\[
p\_harmonic
\]
\[
0.9053312
\]