Austin Statistics

Austin Publishing Group

Open Access Full Text Article

Research Article

Hypotheses Testing in Case-Control Spatial Point Processes

Run Zhuang[1] and Tonglin Zhang[2]*

[1]Department of Biomedical Engineering, Boston University, USA

[2]Department of Statistics, Purdue University, USA

***Corresponding author:** Tonglin Zhang, Department of Statistics, Purdue University, 250 North University Street, West Lafayette, IN 47907-2066, USA, Email: tlzhang@purdue.edu

Received: May 20, 2014; **Accepted:** June 22, 2014; **Published:** June 27, 2014

Abstract

This paper proposes a novel two-sample two-dimensional Kolmogorov-Smirnov type test for the proportionality of intensity functions in case-control spatial point processes. The proposed test statistic is based on the absolute maximum deviation of proportions of points observed in a selected π-system. It shows that the asymptotic null distribution of the test statistic converges in distribution to a two-dimensional functional pinned Brownian sheet, which depends on both the true intensity functions and the shape of the region. However, by carefully selecting the π-system, the asymptotic null distribution may be reduced to the standard Brownian Bridge. Simulation studies show that the proposed test is effective in testing the proportionality in case-control spatial point processes. In an application of the *West Nile* virus study in Nebraska USA, the method shows that the proportionality between the case (i.e. positive to the virus) and the control (i.e. negative to the virus) is violated.

Keywords: Case-control spatial point processes; Two-dimensional functional pinned Brownian sheet; Kolmogorov-Smirnov test; Proportionality

Introduction

A fundamental problem in spatial epidemiology is the understanding of the relationship between risks experienced by humans or animals. A widely used method to address such a problem is to consider a case-control study for a certain risk [1]. In this research, we focus on problems of case-control studies for spatial point processes. In a case-control spatial point process model, data often consist of locations in a specific geographical region which can be classified into two categories: observations from the *case* process (composed of incidence locations of a particular disease) and observations from the *control* process (composed of incidence locations of other diseases). Typically, each observation in the case process presents a positive result to a certain medical test while each observation in the control process presents a negative one. A common task in the analysis of the case-control spatial point patterns is to compare their spatial distributions. For instance, in our data example of Section 5, we are interested in the comparison between the spatial distributions of dead birds on whether or not they were infected by *West Nile* virus. The main interest is to discover whether the spatial distribution of the case process (i.e. positive incidences) and the spatial distribution of the control process (i.e. negative incidences) are the same. The results of the analysis can provide potentially useful information on how the behavior of birds is affected by the infection of the *West Nile* virus.

To make the comparison, we study the relationship between the two spatial point processes by testing whether their intensity functions are proportional. A useful method to facilitate the comparison is to assume that both the case and control processes are inhomogeneous spatial Poisson processes. To compare their distributions, it is sufficient to simply analyze their intensity functions. If the distributions of the case and control processes are the same, then their intensity functions are proportional. This is called the *proportionality* of a case-control study for spatial point processes [2]. A proportional intensity model is derived if the proportionality holds.

In the literature for spatial epidemiology, the proportional intensity model is often used as a baseline assumption for model development. For instance, the spatial distribution of larynx cancer was compared to the spatial distribution of lung cancer around a prespecified location in the Chorley-Ribble area. The proportional intensity model was derived if the spatial distributions of the two cancers around the location were similar [3]. The proportional intensity model has been modified to suggest explanatory variables for the relationship between the two intensity functions [2]. In addition, a statistical model with clustering effects in the case process is also extended from the proportional intensity model [4]. A second-order analysis approach to the proportional intensity model has been also considered [5].

Although it is a useful assumption, the proportionality in case-control spatial point processes may be questionable in real applications. We note that the comparison between two cumulative distribution functions based on the two-sample two-dimensional Kolmogorov-Smirnov (KS) test has been extensively studied in statistical literature [6]. However, little has been done for spatial point processes. In this article, we develop a spatial point process version of the popular two-sample two-dimensional KS test. Our test statistic is constructed in terms of the absolute maximum difference between the observed point proportions from the two processes. A nice property of the proposed method is that the asymptotic null distribution of the test statistic can be derived under only a few weak assumptions. To our best knowledge, this is the first official test that compares the intensity functions between two spatial point processes.

The remainder of the article is organized as follows. In Section 2, we review the necessary background on the two-sample two-dimensional KS test for cumulative distribution functions (CDFs). In

Citation: Zhuang R and Zhang T. Hypotheses Testing in Case-Control Spatial Point Processes. Austin Stat. 2014;1(1): 9.

Section 3, we propose our test for case-control spatial point processes. In Section 4, we present a simulation study to evaluate our testing method. In Section 5, we apply our testing method to the Nebraska *West Nile* data. In Section 6, we conclude this article with a discussion.

Two-Sample Two-dimensional KS Test for CDFs

The KS test was originally proposed for one-sample one-dimensional continuous data [7] and later extended to one-sample multi-dimensional continuous data [8]. The aim of the one-sample KS test is to determine the distribution family of the observed data. Since it is often necessary to compare two distributions, the two-sample KS test is proposed [9]. This method is later extended to a multiple-sample KS test for the comparison of multiple one-dimensional distributions [10]. The idea of the KS test for one-dimensional distributions has later been extended to multi-dimensional cases for the study of astronomical data, which includes the two-sample two-dimensional KS test [11] as well as the two-sample multidimensional-dimensional KS test [12].

Since the focus of this article is to develop a two-sample KS test for spatial point process data, we decide to only review the two-sample two-dimensional KS test. Although the KS test is one of the most important goodness of fit tests based on the empirical distribution functions of random samples, it has not yet been well extended to the multivariate case [13]. The problem is that the asymptotic null distribution of the test statistic is not distribution-free as in the univariate case. Although a method using a simple transformation to make the asymptotic null distribution distribution-free has been proposed [8], this method cannot be used in the two-sample two-dimensional KS test since it involves the unknown true distribution of the observed data in the two-sample problem.

Let X_1 and X_2 be two independent random vectors with CDFs F_1 and F_2 on R^2, respectively, where F_1 and F_2 are unknown. A classical two-sample nonparametric testing problem considers the null hypothesis

$$H_0 : F_1(x) = F_2(x), \forall x \in \mathbb{R}^2$$

against the alternative hypothesis

$$H_1 : F_1(x) \neq F_2(x), \text{ for some } x \in \mathbb{R}^2.$$

This kind of problems arises when given an observed F_1 sample $X_{11},\ldots,X_{1n_1}$ and an observed F_2 sample $X_{21},\ldots,X_{2n_2}$. It must be determined whether the two distributions are equal. The idea of the KS test is to compare the maximum difference between the empirical distributions of sampled data, where a significant difference is concluded if its value is large. Let the empirical distribution of F_1 be

$$\hat{F}_{1,n_1}(x) = \frac{1}{n_1}\sum_{i=1}^{n_1} I_x(X_{1i})$$

and the empirical distribution of F_2 be

$$\hat{F}_{2,n_2}(x) = \frac{1}{n_2}\sum_{i=1}^{n_2} I_x(X_{2i})$$

where $I_x(y) = I_{x_1,x_2}(y_1,y_2)$ with $x=(x_1,x_2)$ and $y=(y_1,y_2)$ for $x,y\in\mathbb{R}^2$ is the indicator function on R^2 which equals one if $y_1 \le x_1$ and $y_2 \le x_2$ and zero otherwise. The two-sample two-dimensional KS statistic, denoted by K_{n_1,n_2}, for testing H_0 against H_1 is defined as

$$K_{n_1,n_2} = \sqrt{n_1 n_2 n_1 + n_2}\ \sup_{x\in R^2}\left|\hat{F}_{1,n_1}(x) - \hat{F}_{2,n_2}(x)\right|, \qquad (1)$$

Where $\sqrt{n_1 n_2/(n_1+n_2)}$ is the standard term used to ensure that the statistic K_{n_1,n_2} converges to a common limiting distribution? The asymptotic null distribution of K_{n_1,n_2} is provided in the following proposition.

Proposition 1

Let $X_{11},\ldots,X_{1n_1}$ and $X_{21},\ldots,X_{2n_2}$ be independently observed from F_1 and F_2, respectively. Then under $F_1 = F_2 = F$,

$$K_{n_1,n_2} \xrightarrow{d} \max_{x\in\mathbb{R}^2} |W_F(x)|, \qquad (2)$$

as $K = n_1 \wedge n_2 = \min(n_1,n_2) \to \infty$, where $W_F(x)$ is the F-functional two-dimensional pinned Brownian sheet, which is a mean, zero Gaussian process on R^2 with the covariance function given by

$$Cov[W_F(x), W_F(x')] = E[W_F(x)W_F(x')] = F(x\wedge x') - F(x)F(x'), \qquad (3)$$

where $x \wedge x' = (x_1\wedge x_1', x_2 \wedge x_2')$ for $x=(x_1,x_2)$ and $x'=(x_1',x_2')$ in $\mathbb{R}^2$.

Proof: Clearly under $F_1 = F_2 = F$ we have $E[\hat{F}_{1,n_1}(x) - \hat{F}_{2,n_2}(x)] = 0$ and

$$\begin{aligned} &Cov[\hat{F}_{1,n_1}(x) - \hat{F}_{2,n_2}(x), \hat{F}_{1,n_1}(x') - \hat{F}_{2,n_2}(x')] \\ &= Cov[\hat{F}_{1,n_1}(x), \hat{F}_{1,n1}(x')] + Cov[\hat{F}_{2,n_2}(x), \hat{F}_{2,n_2}(x')] \\ &= \frac{n_1+n_2}{n_1 n_2}[F(x_1\wedge x_1', x_2\wedge x_2') - F(x_1,x_2)F(x_1',x_2')]. \end{aligned}$$

Therefore, the covariance function of $\sqrt{n_1 n_2/(n_1+n_2)}[\hat{F}_{1,n_1}(x) - \hat{F}_{2,n_2}(x)]$ is the same as the covariance function of $W_F(x)$ for $x\in R^2$. To show the asymptotic distribution of K_{n_1,n_2} given by (2), we need to use the basic theory of the empirical distribution, which includes Theorem 19.4, Theorem 19.5, and the method for the Donsker condition given by Example 19.6 in [14]. First, we consider $\sqrt{n_1}[\hat{F}_{1,n_1}(x) - F(x)]$ as $n_1 \to \infty$. According to the theory of the empirical distribution, it weakly converges to $W_F(x)$ as $n_{1\to\infty}$. A similar conclusion also holds for $\sqrt{n_2}[\hat{F}_{2,n_2}(x) - F(x)]$ as $n_2\to\infty$. Note that these two expressions are independent. We conclude that $\sqrt{n_1 n_2/(n_1+n_2)}[\hat{F}_{1,n_1}(x) - \hat{F}_{2,n_2}(x)]$ weakly converges to $W_F(x)$ as $\min(n_1,n_2)\to\infty$. With the method given by Example 19.6 of [14], we can show that the Donsker condition holds in this case. Then, the conclusion given by (2) is drawn using Theorem 19.5 of [14].

In Proposition, if F is the CDF of the uniform distribution on $[0,1]^2$, then W_F is called the two-dimensional standard pinned Brownian sheet, which is denoted by $W(x)$ for $x\in[0,1]^2$. It is clear that the covariance function of the $W(x)$ is given by $E[W(x)W(x')] = (x_1\wedge x_1')(x_2\wedge x_2') - (x_1x_2)(x_1'x_2')$, $x, x' \in [0,1]^2$.

To conduct the test, one can first directly compute the value of K_{n_1,n_2} defined in (1) and then compare it with the upper tail critical values obtained from its limiting distribution given by (2). As neither the exact nor the approximate distribution of $\max_{x\in R^2}|W_F(x)|$ is known, a Monte Carlo method is often used. Since the distribution of $\max_{x\in R^2}|W_F(x)|$ may depend on F, it is not easy to provide a general list of the critical values for the significance of K_{n_1,n_2}. This issue will be discussed later in Section 4. Using the simulation method, it can be shown that if *F* is the CDF of the uniform distribution on $[0,1]^2$ then the critical values at 1%, 5%, and 10% levels are approximately equal to 1.8656, 1.6522, and 1.4937, respectively. However, this is not enough to carry out a general two-sample KS test for two-dimensional CDFs.

Method

Although much has been done for the comparison between two CDFs, there is little work developed for spatial point processes. In this section, we propose our method, which is modified from the two-sample KS test, for the comparison between two independent spatial point processes. Since the most important issue in a spatial point process is its (first-order) intensity function, we decide to focus on the comparison between the intensity functions of two spatial point processes. If their intensity functions are proportional, then the two spatial point processes will have similar features which indicate that their distributions are affected by the same spatially varying factors. In this article, we consider the simplest case in such a problem: the comparison of intensity functions between the case point process and the control point process in a case-control study, where both case and control point processes can be modeled by inhomogeneous Poisson processes with unknown intensity functions [2].

Spatial point processes

The theory and concept of spatial point processes are well-established, which are available in many textbooks [15-17]. Overall, a spatial point process is defined on a measurable subset in a completely separable metric space. Let the completely separable metric spaces be $\mathbb{R}^2$ and the measurable subset be *S*. Then, a spatial point process *N* is composed of points observed in *S*. Denote *B*(*S*) as the collection of all Borel sets of *S*. Let *N*(*A*) be the number of points in $A\in B(S)$. Then, *N*(*A*) is finite if *A* is bounded. If *N*(*A*) and *N*(*A*’) are independent for any disjoint *A* and *A*’ in *B*(*S*), then *N* is called a spatial Poisson process. If *N* is a spatial Poisson process, then its distribution can be uniquely determined by its intensity function λ(**s**), which is defined by

$$\lambda(s) = \lim_{d(U_s)\to 0}\frac{E[N(U_s)]}{|U_s|} \qquad (4)$$

where U_s is a neighborhood of **s**∈*S*, $|U_s|$ is its Lebesgue measure, and $d(U_S)$ represents the diameter of U_S: $d(U_S)=\max\{d(x,y): x\in U_S, y\in U_S\}$ for a distance function *d*. If *N* is a spatial Poisson process, then *N*(*A*) follows a Poisson distribution with mean $\mu(A)=\int_A \lambda(s)ds$. Further in this section, we propose our methods based on a case-control spatial Poisson process, where both the case and the control processes are modeled by spatial Poisson processes.

The test statistic

Let N_1 and N_2 be two independent spatial Poisson processes on *S* with intensity function $\lambda_1(s)$ and $\lambda_2(s)$, respectively. Then, for any $A\in B(S)$, $N_1(A)$ and $N_2(A)$ are independent Poisson random variables with mean functions $\mu_1(A)=\int_A \lambda_1(s)ds$ and $\mu_2(A)=\int_A \lambda_2(s)ds$, respectively, where both $\lambda_1(s)$ and $\lambda_2(s)$ are positive and continuous. In this article, we focus on testing the null hypothesis of

$$H_0:\lambda_1(s)=\omega\lambda_2(s) \qquad (5)$$

for some ω>0 against the alternative hypothesis of

$$H_1:\lambda_1(s)\neq\omega\lambda_2(s) \qquad (6)$$

for any ω>0, where H_0 implies that $\lambda_1(s)$ and $\lambda_2(s)$ are proportional and H_1 implies that $\lambda_1(s)$ and $\lambda_2(s)$ are not proportional.

Note that H_0 can be interpreted as: there exists an ω>0 such that $E[N_1(A)]=\omega E[N_2(A)]$ for any $A\in B(S)$ and H_1 can be interpreted as: there exists an $A\in B(S)$ such that $E[N_1(A)]\neq\omega E[N_2(A)]$ for any ω>0. Let

$$D_\omega(A)=N_1(A)-\omega N_2(A). \qquad (7)$$

Then, for a given *A*, $D_\omega(A)$ is a function of ω. The null hypothesis is equivalent to that there exists an ω>0 such that

$$\sup_{A\in B(S)}|E[D_\omega(A)]|=0 \qquad (8)$$

and the alternative hypothesis is equivalent to that for any ω>0

$$\sup_{A\in B(S)}|E[D_\omega(A)]>0. \qquad (9)$$

Our test statistic is developed by considering the behavior of $D_\omega(A)$ for all $A\in B(S)$. The basic idea of our approach is formulated in the following theorem.

Theorem 1 *Let S⊆B(S) be a collection of Borel sets in S. If S is a π-system, i.e. S satisfies* $A_1\cap A_2\in S$ *if* $A_1, A_2\in S$*, then a necessary condition for Equation (7) to hold for all A∈B(S) is that there is an ω>0 such that* $E[D_\omega(A)]=0$ *for any A∈S. In addition, if B(S) can be generated by S, then the condition is also sufficient.*

Proof: The necessity can be directly implied by Equation (7). We only need to show the sufficiency. Let $v_\omega(A)=E[D_\omega(A)]$. Then, D_ω is a signed measure for any ω>0. According to the Hahn Decomposition ([18], P 420), we can find S^+ and S^- with $s^+\cap s^-=\phi$ and $s^+\cup s^-=s$ such that v_ω can be almost surely uniquely decomposed into $v_\omega=v_\omega^+-v_\omega^-$ with $v_\omega(A)=v_\omega^+(A)-v_\omega^-(A)$ for any $A\in B(S)$, where $v_\omega^+(A)=v_\omega(A\cap S^+)$ and $v_\omega^-=-v_\omega(A\cap S^-)$ are two nonnegative σ-finite measures on *S*. Let ω_0 be the true value of ω such that $v_\omega(A)=0$ for all *A*∈*S*. Then, $v_{\omega_0}^+$ and $v_{\omega_0}^-$ agree on *S*. According to the theorem of the π-λ system which says that if two measures agree on a π-system then they also agree on the σ-algebra of the π-system (e.g. Theorem 3.3 in [18], P 42), we conclude that $v_{\omega_0}^+$ and $v_{\omega_0}^-$ agree on σ(*S*)=*B*(*S*). This is enough to conclude the sufficiency.

It is clear from Theorem that Equation (7) is only necessary to be considered in a special π-system *S*⊆*B*(*S*), which implies that we only need to consider

$$\sup_{A\in S}|E[D_\omega(A)]|=0. \qquad (10)$$

It can be seen from Theorem that H_0 is rejected if Equation (10) is violated. However, if Equation (10) holds, then H_0 is accepted only when (*S*) can be generated by *S*. Note that under H_0 a straightforward estimator of ω is

$$\hat{\omega}=\frac{N_1(S)}{N_2(S)}. \qquad (11)$$

Then,

$$D_{\hat{\omega}}(A)=N_1(S)\left[\frac{N_1(A)}{N_1(S)}-\frac{N_2(A)}{N_2(S)}\right].$$

Using the above in (10), we derive our test statistic as

$$T=\sqrt{N_1(S)N_2(S)N_1(S)+N_2(S)}\sup_{A\in S}\left|\frac{N_1(A)}{N_1(S)}-\frac{N_2(A)}{N_2(S)}\right|, \qquad (12)$$

where *S* is a collection of a π-system in *B*(*S*), and H_0 is rejected if *T* is large. The *p*-value of *T* can be derived from the distribution of the maximum of an *F*-functional pinned Brownian sheet, where the function *F* can be defined using the following theorem. The conditions in the asymptotics considered in the theorem can be interpreted as the expected number of points in the two processes, i.e. $E[N_1(S)]$ and $E[N_2(S)]$, approaches infinity but the proportion of expected points in subregions of *S* does not vary.

Theorem 2 *Let the π-system*

$$S_G=\{A_X\in B(S): A_X=G^{-1}([0,X_1]\times[0,X_2]), X=(X_1,X_2)\in[0,1]^2\} \qquad (13)$$

be generated by a measurable function G from S to $[0,1]^2$. *Denote T as* T_G *if the π-system in T is given by* S_G. *Define* $F_1(X)=\mu_1(A)_X/\mu_1(S)$ *and* $F_2(X)=\mu_2(A)_X/\mu_2(S)$. *Assume for any* $x\in R^2$ *the two functions* $F_1(x)$ *and* $F_2(x)$ *do not vary as* $k=\min(E[N_1(S)],E[N_2(S)])\to\infty$. *If there exists a positive ω such that* $\lambda_1(s)=\omega\lambda_2(s)$ *(which is also* $F_1=F_2=F$ *), then* $T_G\xrightarrow{D}\sup_{x\in[0,1]^2}|W_F(x)|$ *as* $k\to\infty$.

Proof: Note that the spatial Poisson processes N_1 and N_2 can be interpreted as points being independently observed from *S* with CDFs F_1 and F_2, respectively. Then, $N_1(A_X)|N_1(S)\sim Bin(N_1(S),F_1(x))$ and $N_2(A_X)|N_2(S)\sim Bin(N_2(S),F_2(x))$. Denote $k_1=E[N_1(S)]$, $k_2=E[N_2(S)]$ and Ā as the complementary set of *A*. For any $x,x'\in[0,1]^2$, define

$$Z=Z_{x,x'}=\frac{1}{\kappa_1+\kappa_2}(N_1(A_x\cap A_{x'}),N_1(A_x\cap\overline{A}_{x'}),N_1(\overline{A}_x\cap A_{x'}),N_1(\overline{A}_x\cap\overline{A}_{x'}),$$

$$N_2(A_x\cap A_{x'}),N_2(A_x\cap\overline{A}_{x'}),N_2(\overline{A}_x\cap A_{x'}),N_2(\overline{A}_x\cap\overline{A}_{x'}))^t$$

and

$$V=V_{x,x'}=\frac{1}{\kappa_1+\kappa_2}(\mu_1(A_x\cap A_{x'}),\mu_1(A_x\cap\overline{A}_{x'}),\mu_1(\overline{A}_x\cap A_{x'}),\mu_1(\overline{A}_x\cap\overline{A}_{x'}),$$

$$\mu_2(A_x\cap A_{x'}),\mu_2(A_x\cap\overline{A}_{x'}),\mu_2(\overline{A}_x\cap A_{x'}),\mu_2(\overline{A}_x\cap\overline{A}_{x'}))^t.$$

Then, *Z* is an eight-dimensional independent Poisson random vector with the mean vector given by ν. According to the central limit theorem for independent Poisson random variables, we have

$$\sqrt{k_1+k_2}(Z-v)\xrightarrow{d}N\left(0,\frac{diag(v)}{k_1+k_2}\right),$$

and $k=\min(k_1,k_2)$. Let $h(z)=(h_1(z),h_2(z))^t$ with $h_1(z)=(z_1+z_2)/(z_1+z_2+z_3+z_4)-(z_5+z_6)/(z_5+z_6+z_7+z_8)$ and $h_2(z)=(z_1+z_3)/(z_1+z_2+z_3+z_4)-(z_5+z_7)/(z_5+z_6+z_7+z_8)$, for $z=(z_1,\dots,z_8)\in R^8$. Then, $h_1(Z)=N_1(A_X)/N_1(S)-N_2(A_X)/N_2(S)$, $h_2(Z)=N_1(A_{X'})/N_1(S)-N_2(A_{X'})/N_2(S)$, $h_1(v)=F_1(x)-F_2(x)$, and $h_2(v)=F_1(x')-F_2(x')$. Let $\dot h_1$ and $\dot h_2$ be the gradients of h_1 and h_2, respectively. Then,

$$\dot h_1(v)=(k_1+k_2)\left(\frac{\mu_1(\overline{A}_x)}{\mu_1^2(S)},\frac{\mu_1(\overline{A}_x)}{\mu_1^2(S)},-\frac{\mu_1(A_x)}{\mu_1^2(S)},-\frac{\mu_1(A_x)}{\mu_1^2(S)},-\frac{\mu_2(\overline{A}_x)}{\mu_2^2(S)},-\frac{\mu_2(\overline{A}_x)}{\mu_2^2(S)},\frac{\mu_2(A_x)}{\mu_2^2(S)},\frac{\mu_2(A_x)}{\mu_2^2(S)}\right)^t$$

and

$$\dot h_2(v)=(k_1+k_2)\left(\frac{\mu_1(\overline{A}_{x'})}{\mu_1^2(S)},-\frac{\mu_1(A_{x'})}{\mu_1^2(S)},\frac{\mu_1(\overline{A}_{x'})}{\mu_1^2(S)},-\frac{\mu_1(A_{x'})}{\mu_1^2(S)},-\frac{\mu_2(\overline{A}_{x'})}{\mu_2^2(S)},\frac{\mu_2(A_{x'})}{\mu_2^2(S)},\frac{\mu_2(\overline{A}_{x'})}{\mu_2^2(S)},\frac{\mu_2(A_{x'})}{\mu_2^2(S)}\right)^t.$$

If $\lambda_1(s)$ and $\lambda_2(s)$ are proportional, then $h_1(v)=h_2(v)=0$ and

$$\begin{pmatrix}\dot h_1(v)\\ \dot h_2(v)\end{pmatrix}\frac{diag(v)}{\kappa_1+\kappa_2}\begin{pmatrix}h_1(v)\\ \dot h_2(v)\end{pmatrix}^t=\frac{(\kappa_1+\kappa_2)^2}{\kappa_1\kappa_2}\begin{pmatrix}F(x)[1-F(x)] & F(x\wedge x')-F(x)F(x')\\ F(x\wedge x')-F(x)F(x') & F(x')[1-F(x')]\end{pmatrix}.$$

Using the Delta Method with the expressions of the gradients of $h_1(z)$ and $h_2(z)$ given above, we have

$$\sqrt{\frac{\kappa_1\kappa_2}{\kappa_1+\kappa_2}}\begin{pmatrix}N_1(A_x)/N_1(S)-N_2(A_x)/N_2(S)\\ N_1(A_{x'})/N_1(S)-N_2(A_{x'})/N_2(S)\end{pmatrix}$$

$$\xrightarrow{d}N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}F(x)[1-F(x)] & F(x\wedge x')-F(x)F(x')\\ F(x\wedge x')-F(x)F(x') & F(x')[1-F(x')]\end{pmatrix}\right)$$

as $k\to\infty$. Since the above holds for any pair of $x,x'\in[0,1]^2$, under $F_1=F_2=F$ the covariance function of $N_1(A_X)/N_1(S)-N_2(A_X)/N_2(S)$ is the same as the covariance function of $W_F(x)$, which implies the conclusion of the theorem by the theory of empirical distributions that was already used in the proof of Proposition 1.

According to Theorem, the *p*-value of *T* can be approximately derived by the distribution of $\sup_{x\in[0,1]^2}|W_F(x)|$. Since there is no closed form formula of such a distribution, a Monte Carlo method is used. This issue will be discussed in our simulation study in Section 4.

Practical guidelines

To calculate the test statistic *T*, it is important to choose the pre-selected function *G* from *S* to $[0,1]^2$ to determine the π-system S_G. Generally, *G* is defined via a continuous bivariate function. Assume that proportionality holds and denote $\lambda_1(s)=\lambda_2(s)=\lambda(s)$. The basic idea can be derived by considering the special case in which *S* is a rectangular region given by *S*=[0,*a*]×[0,*b*] for *a*,*b*>0. In this case, a natural choice of *G* is

$$G_{a,b}(s)=\frac{s_1s_2}{ab} \qquad (14)$$

for $s=(s_1,s_2)\in[0,a]\times[0,b]$. The corresponding *F*-functional pinned Brownian sheet on $[0,1]^2$ is derived if we choose *F* as

$$F(x)=F_{a,b}(x)=\frac{\int_{[0,x_1a]\times[0,x_2b]}\lambda(s)ds}{\int_{[0,a]\times[0,b]}\lambda(s)ds}, x=(x_1,x_2)\in[0,1]^2, \qquad (15)$$

which can be used to compute the *p*-value of $T_{G_{a,b}}$ using the distribution of $\sup_{x\in[0,1]^2}|W_{F_{a,b}}(x)|$. For an arbitrary region *S* in R^2, we can define

$$G_{s_0}(s)=\left(\frac{|s_1-s_{01}|}{\max_{s'\in S}|s_1'-s_{01}|},\frac{|s_2-s_{02}|}{\max_{s'\in S}|s_2'-s_{02}|}\right), s=(s_1,s_2)\in S, \qquad (16)$$

where $s_0=(s_{01},s_{02})$ is a pre-selected point in *S*. The corresponding *F*-functional pinned Brownian sheet on $[0,1]^2$ is derived if we choose *F* as

$$F(x)=F_{s_0}(x)=\frac{\int_{G^{-1}([0,x_1]\times[0,x_2])}\lambda(s)ds}{\int_S\lambda(s)ds}. \qquad (17)$$

Since the distribution of $\sup_{x\in[0,1]^2}|W_F(x)|$ with $F=F_{a,b}$ in (15) or $F=F_{s_0}$ in (16) depends on *F*, it is generally impossible to provide a general numerical table for $T_{G_{a,b}}$ or $T_{G_{s_0}}$, which implies that the critical values of the test should be provided by a Monte Carlo method in every application. In order to avoid this difficulty, we consider a simplified choice of *G* as

$$G_{s_0,0}(s)=\frac{\|s-s_0\|}{\max_{s'\in S}\|s'-s_0\|}. \qquad (18)$$

It can be seen that such an *F* can make $W_F(x)$ to be the standard Brownian bridge on [0,1]. The Taylor expansion of the distribution of the absolute maximum of the standard Brownian bridge on [0,1] is well-known and available in many textbooks [14]. According to the Taylor expansion, we can approximately compute the p-value of $T_{G_{s_0,0}}$ by

$$P(T_{G_{s_0,0}} \ge c) \approx 2\sum_{k=1}^{\infty}(-1)^{k+1}e^{-2k^2c^2}. \quad (19)$$

If (19) is used, the critical values at 10%, 5%, and 1% levels are approximately equal to 1.2239, 1.3581, and 1.6277, respectively (Figure 1).

Simulation

We considered both a rectangular region and an arbitrary region in our simulations. We chose these regions because we wanted to evaluate the efficiency and accuracy of our test when G is given by (14), (16), and (18), respectively. We considered the rectangular region because we wanted to know how the test was influenced by the intensity functions of the two spatial Poisson processes and considered an arbitrary region because we wanted to know how the test was influenced by the shape of the region, both of which are important in applications.

Rectangular Region

We selected a squared region S=[0,20]×[0,20] in the study of rectangular regions (Figure1). We simulated realizations from two independent spatial Poisson processes, denoted by N_1 and N_2, respectively, on S. We chose the intensity function of N_1 equafig: rectangular to

$$\lambda_1(s) = \frac{k}{2\pi\sqrt{1-\rho^2}} e^{-\frac{(s_1-10)^2-2\rho(s_1-10)(s_2-10)+(s_2-10)^2}{2(1-\rho^2)}},$$

and the intensity function of N_2 equal to

$$\lambda_2(s) = \frac{k}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-\frac{(s_1-10)^2-2\rho(s_1-10)(s_2-10)+(s_2-10)^2}{2\sigma^2(1-\rho^2)}},$$

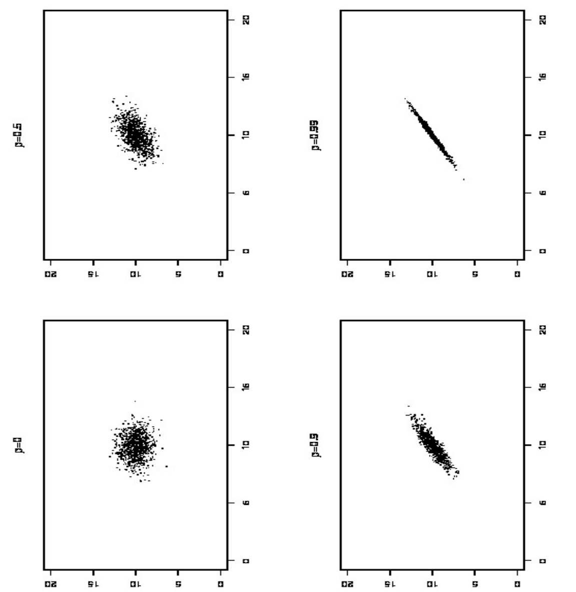

Figure 1: Realizations of points in the case process in the [0,20]×[0,20] region when κ=1000 for ρ equal to 0, 0.5, 0.9, and 0.99, respectively.

for $s = (s_1, s_2) \in [0,10]^2$, where κ was a constant. It can be seen from above that both intensity functions were proportional to the probability density function (PDF) of the bivariate normal distribution restricted on S with mean (10,10), variance (σ^2, σ^2), and correlation p, where the variance of N_1 was always equal to 1 but the variance of N_2 was subjected to change. From the properties of the bivariate normal, we had $E[N_1(S)] \approx k$ and $E[N_2(S)] \approx k$ in this setting and the two intensity functions were proportional iff $\sigma^2 = 1$ in $\lambda_2(s)$.

We considered $G = G_{20,20}(s)$ with $G_{20,20}(s)$ given by (14) and $G = G_{S_0,0}(s)$ for $s_0 = (0,0)$ with $G_{s_0,0}(s)$ given by (18). Therefore, we had two test statistics in the simulation: $T_{G_{20,20}}$ and $T_{G_{(0,0),0}}$. Based on Theorem, the asymptotic null distribution of T in the former case might be determined by the distribution of $\sup_{x\in[0,1]^2}|W_F(x)|$ with F given by (15). The asymptotic null distribution of T in the latter case might be determined by a Taylor expansion given by (19). Therefore, in addition to the type I an error probability and power functions, the accuracy of the asymptotic null distributions provided by Theorem was also an important issue.

Our simulation studies contained two parts. In the first part, we evaluated the accuracy of the asymptotic null distributions of $T_{G_{20,20}}$ and $T_{G_{(0,0),0}}$ under the null hypothesis provided by Theorem. We carried out a simulation study to evaluate the accuracy of the asymptotic null distributions. To do so, we chose κ=10000 and simulated N_1 and N_2 with σ=1 in $\lambda_2(s)$. Because the asymptotic null distribution of $T_{G_{20,20}}$ might depend on F, we computed its 10%, 5%, and 1% upper quantiles in the simulation. To compare, we also computed the 10%, 5%, and 1% upper quantiles of $T_{G_{(0,0),0}}$. As Theorem needs κ to approach infinity, we considered a simulation study with 10000 replications when κ was large (i.e. κ=10000). The result is displayed in Table1. It shows that the critical values of $T_{G_{20,20}}$ might depend on p but the critical values of $T_{G_{(0,0),0}}$ might not. Therefore, the latter case would be more reliable to be used in applications.

In the second part, we evaluated the performance of the type I error probabilities and power functions of $T_{G_{20,20}}$ and $T_{G_{(0,0),0}}$ with p-values derived from their asymptotic null distributions. We used a 0.05 significance level in the test. We noted that the asymptotic null distribution of $T_{G_{20,20}}$ only deviated from the distribution of $\sup_{x\in[0,1]^2}|W(x)|$ when was close to one. We focused on the case when p was much lower than one, which included two cases (i.e. p=0.0 and p=0.5) in our simulation studies. Therefore, we assumed that the asymptotic null distribution of $T_{G_{20,20}}$ could be approximately derived using the distribution of $\sup_{x\in[0,1]^2}|W(x)|$, which induced significance of the test if $T_{G_{20,20}} > 1.6522$. Because the asymptotic null distribution of $T_{G_{(0,0),0}}$ did not depend on p, significance of the test was concluded if $T_{G_{(0,0),0}} > 1.3581$ in the latter case. We considered σ equal to 1.0, 0.9, and 0.8 in $\lambda_2(s)$ and simulated 1000 realizations using each selected parameter. The result is display Table 2. It shows that the type I error probabilities (when σ=1.0) were not heavily biased in either cases. The power functions (when σ=0.9 or σ=0.8) of $T_{G_{20,20}}$ were slightly less than the power function of $T_{G_{(0,0),0}}$, which indicated that the latter case was more powerful than the former case in the setting that we had considered in our simulations.

An arbitrary region

In order to consider the influence of shapes of regions, we

Table 1: Upper quantiles of the asymptotic null distributions (i.e. σ=1.0) of T derived from simulations with 10^4 replications in $[0,20]^2$ for $T_{G_{(0,0),0}}$ in (14) and $T_{G_{(0,0),0}(s)}$ in (18), respectively.

	$T_{G_{20,20}}$			$T_{G_{(0,0),0}}$		
p	10%	5%	1%	10%	5%	1%
0	1.4990	1.6334	1.9021	1.1809	1.3293	1.6122
0.5	1.4283	1.5698	1.8384	1.1780	1.3285	1.5968
0.9	1.3294	1.4637	1.7254	1.1808	1.3223	1.5987
0.99	1.2586	1.4071	1.6900	1.1816	1.3223	1.6052

Table 2: Type I error probabilities (i.e. σ=1.0) and power functions (i.e. σ=0.9 and σ=0.8) of T at 0.05 significance levels derived from simulations with 1000 replications in $[0,20]^2$ for $G=G_{20,20}$ in (14) and $G=G_{S_0,0}(s)$ with in (18), respectively.

		$T_{G_{20,20}}$ for different σ			$T_{G_{(0,0),0}}$ for different σ		
p	κ	1.0	0.9	0.8	1.0	0.9	0.8
0.0	1000	0.035	0.329	0.937	0.041	0.442	0.993
	2000	0.048	0.652	0.997	0.038	0.780	1.000
0.5	1000	0.031	0.203	0.854	0.036	0.296	0.955
	2000	0.039	0.459	0.990	0.049	0.619	1.000

designed the study region as

$$S_a=\{(x_1,x_2)\in[0,1]^2:|x_1-x_2|\le a, 0\le a\le 1\},$$

which was treated as an irregular region (Figure 2). It was approximately expanding along the diagonal line in $[0,1]^2$. It can be easily shown that the area of S_a was equal to $|S_a|=1-(1-a)^2$ and $S_a=[0,1]^2$ if a=1. We chose the intensity function of N_1 equal to

$$\lambda_1(s)=\frac{k}{1-(1-a)^2}$$

and the intensity function of N_2 equal to

$$\lambda_2(s)=\frac{\kappa\Gamma(2\beta)}{C\Gamma^2(\beta)}\left(\frac{s_1+s_2}{2}\right)^{\beta}\left(1-\frac{s_1+s_2}{2}\right)^{\beta},$$

for $s=(s_1,s_2)\in S_a$. Then, points of N_1 were uniformly distributed on S_a with $E[N_1(S_a)]=k$. Points of N_2 were modified from a Beta distribution on the parallel line to the diagonal of $[0,1]^2$. The intensity functions of N_1 and N_2 were proportional iff β=1. In the simulation, we always selected the value of C such that $E[N_2(S_a)]=k$.

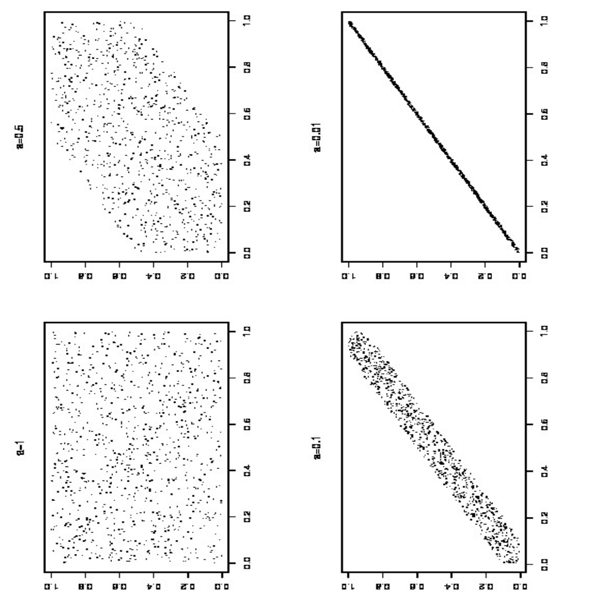

Figure 2: Realizations of points in the case process in the S_a region when κ=1000 for a equal to 1, 0.5, 0.1, and 0.01, respectively.

We considered $T_{G_{(0,0)}}$ and $T_{G_{(0,0),0}}$ given by (16) and (18), respectively. Similar to the discussion in Section4.1, the asymptotic null distribution of $T_{G_{(0,0)}}$ might depend on the shape of the region but the asymptotic null distribution of $T_{G_{(0,0),0}}$ might not. Therefore, we evaluated the accuracy of the asymptotic null distribution, the type I error probabilities, and the power functions of the two test statistics. To evaluate the accuracy of the asymptotic null distribution, we chose κ=10000 in a simulation study with 10000 replications (Table 3). The result showed that the asymptotic null distribution of $T_{G_{(0,0)}}$ might depend on the shape of the region but the asymptotic null distribution of $T_{G_{(0,0),0}}$ might not. Therefore, the latter case would be more reliable in applications.

In addition, we evaluated the type I error probabilities and the power functions of $T_{G_{(0,0)}}$ and $T_{G_{(0,0),0}}$ with p-values derived from their asymptotic null distributions, respectively. We noted that the asymptotic null distribution of $T_{G_{(0,0)}}$ was only heavily biased from the distribution of $\sup_{x\in[0,1]^2}|W(x)|$ when a was close to one. We focused on the case when a was not close to one. These included two cases (i.e. a=1.0 and a=0.5) in our simulation studies. Similar to before, we concluded significance if $T_{G_{(0,0)}}>1.6522$ or $T_{G_{(0,0),0}}>1.3581$ based on the two test statistics. We considered β equal to 1, 2, and 3 in $\lambda_2(s)$ and simulated 1000 realizations using each selected parameters (Table 4). The result showed that the type I error probabilities (when β=1) were not heavily biased in both cases and the power functions (when β=2 and β=3) of $T_{G_{(0,0)}}$ were slightly less than the power function of the case for $T_{G_{(0,0),0}}$. This indicated that the latter case was more powerful than the former case in the setting of our simulations.

In summary, the asymptotic null distribution of T generally depends on the intensity functions of N_1 and N_2 as well as the shape of the study region, but it may not if the π-system in Theorem is carefully considered. As long as the π-system S is selected, it is enough to reject the null hypothesis if the test is significant. However, the

Table 3: Upper quantiles of the asymptotic null distributions (i.e. β=1.0) of $T_{G_{(0,0)}}$ given by (16) and $T_{G_{(0,0),0}}$ given by (18) derived from Monte Carlo simulations with 10^4 replications in S_a.

	$T_{G_{(0,0)}}$			$T_{G_{(0,0),0}}$		
a	10%	5%	1%	10%	5%	1%
1	1.5052	1.6341	1.8998	1.1764	1.3068	1.5909
0.5	1.4545	1.5849	1.8390	1.1760	1.3166	1.5757
0.1	1.2918	1.4297	1.7006	1.1983	1.3386	1.5898
0.01	1.2134	1.3428	1.6247	1.1890	1.3143	1.5917

Table 4: Type I error probabilities (i.e. β=1.0) and power functions (i.e. β=1.5 and σ=2.0) of $T_{G_{(0,0)}}$ and $T_{G_{(0,0),0}}$ at 0.05 significance levels derived from simulations with 1000 replications in G_a.

		$T_{G_{(0,0)}}$ for different β			$T_{G_{(0,0),0}}$ for different β		
β	κ	1	2	3	1	2	3
1.0	1000	0.040	0.291	0.915	0.046	0.341	0.952
	2000	0.042	0.743	1.000	0.048	0.812	1.000
0.5	1000	0.032	0.576	0.998	0.035	0.681	0.999
	2000	0.037	0.969	1.000	0.042	0.983	1.000

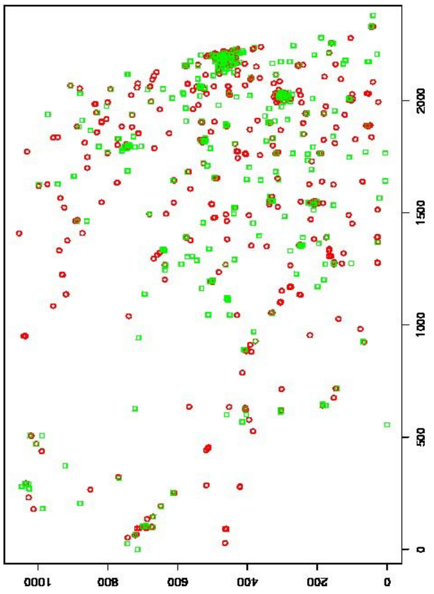

Figure 3: Dead bird locations with respect to the positive (circle and red) and negative (square and green) to the *West Nile* virus in the *Nebraska West Nile* data, where the units in the horizontal and vertical axes are given by kilometers.

reverse may not be true since $B(S)$ may not be generated by S.

Application

We applied our proposed method to the *Nebraska West Nile* data. The data contained the locations of dead birds from 2002 to 2013 in Nebraska, USA, which might be related to an infection of the *West Nile* virus. In Nebraska, *West Nile* virus surveillance effects focus on the late summer and early fall months. Since *West Nile* virus was discovered in US in 1999, the virus has been detected in over 300 species of dead birds. The virus was transmitted to birds through the bite of infected mosquitoes, where mosquitoes became infected by biting infected birds. Some birds might have become infected after consumption of sick or dead birds that were already infected by *West Nile* virus. Some infected birds were known to get sick and died from the infection. Reporting and testing of dead birds was one way to test for the presence of *West Nile* virus. The *West Nile* virus might also be transmitted to humans. In Nebraska, the most serious year of *West Nile* virus for humans was 2003, which contributed to around 60% of infections and 50% deaths in the whole twelve study period. Therefore, we focused on the analysis of the spatial patterns of dead birds in Nebraska in 2003.

Based on the results of testing of *West Nile* virus for dead birds in 2003, there were 576 positive occurrences and 454 negative occurrences (Figure 3). We assumed that the locations of positive dead birds were observed from the case process and the locations of the negative dead birds were observed from the control process. We focused the test on whether the intensity (i.e. $\lambda_1(s)$) of the case process was proportional to the intensity (i.e. $\lambda_2(s)$) of the control process. To apply our test, we used $s_0=(0,0)$ in G_{s_0} and $G_{s_0,0}$ in Equations (16) and (18) and derived $G_{(0,0)}$ and $G_{(0,0),0}$, respectively. When $G_{(0,0)}$ was used, the value of $T_{G_{(0,0)}}$ was 1.7286. According to the distribution of $\sup_{x\in[0,1]^2}|W(x)|$, the p-value was 0.0282. When $G_{(0,0),0}$ was used, the value of $T_{G_{(0,0),0}}$ was 1.7075. According to the distribution of $\sup_{x\in[0,1]}|W(x)|$, the p-value was 0.0059. Both were significant at the 0.05 significance level. Therefore, we concluded that $\lambda_1(s)$ and $\lambda_2(s)$ were not proportional, which indicated that their spatial patterns were not similar. Note that we had κ_1 was around 576 and κ_2 was around 454. These were large enough for us to apply the asymptotic null distributions to compute the p-values of our test statistics.

After we had concluded that the intensity functions between the case process and the control process were not proportional, the next interest was to discover their difference. We considered the ratio of the two intensity functions. We used the concept of a spatial cluster model [3], which described the variation of the ratio via a logistic regression model as

$$\log\frac{\lambda_1(s)}{\lambda_2(s)}=\alpha+\varrho f(s-\gamma;\Theta), \qquad (20)$$

where α, p, Θ, and γ are unknown parameters. In Equation (20), the intercept α reflected the logarithm of the ratio of the overall number of case events relative to the overall number of control events. The function $f(.)$ described the changes of the ratio (i.e. the relative risks) with position to $\gamma = (\gamma_1, \gamma_2)$ in the study region. Although the function $f(.)$ in Equation (20) could be very general, we adopted the usual bivariate Gaussian density functions as

$$f(s;\Theta) = e^{-\frac{[\Theta_1(s_1-\gamma_1)]^2}{2} - \frac{[\Theta_2(s_2-\gamma_2)]^2}{2}}, \Theta = (\Theta_1, \Theta_2) \in R^{+2}. \quad (21)$$

In Equation (21), $1/\Theta_1$ and $1/\Theta_2$ were scale parameters. We considered maximum likelihood estimation for parameters in Model (20). It showed that the best Θ_1 was close to zero. Therefore, we revised the model as

$$f_{rev}(s;\Theta) = e^{-\frac{[\Theta_2(s_2-\gamma_2)]^2}{2}}, \Theta_2 \in R^+, \quad (22)$$

which resulted the parameters in Model (20) to be α, p, Θ_2, and γ_2. The MLE of these parameters were ^α=0.7357, ^p =−0.7480, $1/\Theta_2 = 234.7$, and $\gamma_2 = 540.6(km)$.

According to the MLE of the parameters in Model (20), we concluded that the ratio between $\lambda_1(s)$ and $\lambda_2(s)$ was almost identical if a point moved horizontally but it increased as the point moved either to the south or to the north from the central line of the state (i.e. y=540.6km). The ratio attained its minimum value (0.9860) at the y=540.6km line and gradually attained its maximum at south (1.9798) or its maximum at north (1.9490). Therefore, we expected to see relative more *West Nile* birds in the north part or south part in the state.

Discussion

We proposed a two-sample KS test to assess the proportionality between the intensity functions of two independent Poisson processes. Our method is modified from the classic two-sample KS test for the equality of CDFs. The difference between the construction of the usual two-sample KS test for CDFs and our two-sample KS test for intensity functions is that the π-system used in the KS test for CDFs has a naturally defined π-system, but such a π-system does not exist in the KS test for intensity functions. Therefore, the π-system must be selected. There are many different ways to select the π-systems, which may result in many different versions of our test statistic. According to our theoretical conclusions that we have presented in Section 3, it is enough to conclude the alternative hypothesis if a test is significant. However, if the test is insignificant, then we can only conclude the null hypothesis if the σ-algebra of $\mathbb{R}^2$ can be generated by the selected π-system. Therefore, an insignificant test may be not enough to conclude the acceptance of the null hypothesis.

After the π-system has been selected, we have shown that the asymptotic null distribution of our test statistic either weakly converges to the maximum absolute value of an F-functional pinned Brownian sheen on $\mathbb{R}^2$ or weakly converges to the maximum absolute value of the standard Brownian bridge on [0,1]. Although it has been known that the distribution of the previous one may depend on F, we have numerically shown that such a distribution can be approximated by the maximum absolute value of the standard pinned Brownian sheet on $\mathbb{R}^2$ provided that F is not very close to a degenerate case. Therefore, using the p-value derived from the absolute maximum of the two-dimensional standard Brownian sheet may not be seriously biased in most cases of applications. This method is adopted in our application section. For the latter case, the asymptotic null distribution of our test statistic is uniquely determined. Since the π-system used in the construction of the test statistic cannot generate the σ-algebra of $\mathbb{R}^2$, an insignificant test is not enough to conclude the acceptance of the null hypothesis. Therefore, the interpretation of the test should be carefully addressed.

As long as the rejection of the null hypothesis is concluded, the next step is to compare the difference between the two intensity functions. A powerful tool is to use the spatial cluster modeling technique, which specifies a spatial cluster term in a statistical model. However, the significance of the spatial cluster term in the model is not easily understood since this involves a famous statistical testing problem which concerns hypothesis testing when some of the parameters are only present in the alternative hypothesis [19]. For instance, in the method that we have used in Section 5, the parameters γ and Θ is only present when $p \neq 0$. The usual χ^2-test for the significance of the second term in Model (20) is invalid. In this case, we may use the p-value derived from the two-sample two-dimensional KS test for significance.

Acknowledgement

The authors gratefully acknowledge the suggestive comments from editor, associate editor, and referees, which led to a much improved article.

References

1. Jones MP, O'Gorman TW, Lemke JH, Woolson RF. A Monte Carlo investigation of homogeneity tests of the odds ratio under various sample size configurations. Biometrics. 1989; 45: 171-181.
2. Diggle PJ, Gómez-Rubio V, Brown PE, Chetwynd AG, Gooding S. Second-order analysis of inhomogeneous spatial point processes using case-control data. Biometrics. 2007; 63: 550-557.
3. Diggle PJ. A point process modeling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. Journal of the Royal Statistical Society B. 1990; 153: 349–362.
4. Henrys PA, Brown PE. Inference for clustered inhomogeneous spatial point processes. Biometrics. 2009; 65: 423-430.
5. Guan Y, Waagepetersen R, Beale CM. Second-order analysis of inhomogeneous spatial point processes with proportional intensity functions. Journal of the American Statistical Association. 2008; 103: 769-777.
6. Lopes RHC, Reid I, Hobson PR. The two-dimensional Kolmogorov-Smirnov test. In XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research, Amsterdam. The Netherlands. 2007.
7. Smirnov N. Table for estimating the goodness of fit of empirical distributions. Annals of Mathematical Statistics. 1948; 19; 279-281.
8. Rosenblatt M. Remarks on a multivariate transformation. Annals of Mathematical Statistics. 1952; 23: 470-472.
9. Hodges J. The significance probability of the Smirnov two-sample test. Matematik. 1957; 12: 469-486.
10. Kiefer J. K-sample analogues of the Kolmogorov-Smirnov and Cramér-V. Mises test. The Annals of Mathematical Statistics. 1959; 30: 420-447.
11. Peacock JA. Two-dimensional goodness-of-fit testing in astronomy. Monthly Notices of Rotal Astronomical Society. 1983; 202: 615–627.
12. Fasano G, Franceschini, A. A multidimensional version of the Kolmogorov-

Smirnov test. Monthly Notices of Royal Astronomical Society. 1987; 225: 155-170.

13. Justel A, Peña D, Zamar R. A multivariate Kolmogorov-Smirnov test of goodness of fit. Statistics and Probability Letters. 1997; 35: 251-259.

14. Van der Vaart AW. Asymptotic Statistics. Cambridge UK: Cambridge University Press. 1998.

15. Daley DJ, Vere-Jones D. An Introduction to Theory of Point Processes: Volume I: Elementary Theory and Methods, Second Edition. New York: Springer. 2003.

16. Diggle PJ, Ribeiro JPJ, Christensen O. An introduction to model-based geostatistics. In Spatial Statistics and Computational Methods. Moller J, editor. New York: Spring. 2003; 173: 43-86.

17. Karr AF. Point Processes and Their Statistical Inference 2nd Edition. New York: Marcel Dekker Inc. 1991.

18. Billingsley P. Probability and Measure. New York: John Wiley & Sons. 1995.

19. Zheng G, Chen Z. Comparison of maximum statistics for hypothesis testing when a nuisance parameter is present only under the alternative. Biometrics. 2005; 61: 254-258.

Austin Stat - Volume 1 Issue 1 - 2014
Submit your Manuscript | www.austinpublishinggroup.com

Citation: Zhuang R and Zhang T. Hypotheses Testing in Case-Control Spatial Point Processes. Austin Stat. 2014;1(1): 9.