Bernoulli **23**(1), 2017, 89–109
DOI: 10.3150/15-BEJ727

Asymptotic properties of spatial scan statistics under the alternative hypothesis

TONGLIN ZHANG[1] and GE LIN[2]

[1]*Department of Statistics, Purdue University, 250 North University Street, West Lafayette, IN 47907-2066, USA. E-mail: tlzhang@purdue.edu*
[2]*Department of Environmental and Occupational Health, School of Community Health Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA. E-mail: ge.kan@unlv.edu*

A common challenge for most spatial cluster detection methods is the lack of asymptotic properties to support their validity. As the spatial scan test is the most often used cluster detection method, we investigate two important properties in the method: the consistency and asymptotic local efficiency. We address the consistency by showing that the detected cluster converges to the true cluster in probability. We address the asymptotic local efficiency by showing that the spatial scan statistic asymptotically converges to the square of the maximum of a Gaussian random field, where the mean and covariance functions of the Gaussian random field depends on a function of at-risk population within and outside of the cluster. These conclusions, which are also supported by simulation and case studies, make it practical to precisely detect and characterize a spatial cluster.

Keywords: asymptotic distribution; clusters; converges in probability; Gaussian random field; spatial scan statistics

1. Introduction

Spatial cluster detection is a critical part of disease surveillance, and many cluster detection methods have been developed in the last two decades (see a thorough review by [16] as well as [2,5,7,11,13]). However, most cluster detection methods share the problem of how to ascertain and characterize a spatial cluster. R. Fisher noted that in agriculture experiments "patches in close proximity are commonly more alike… than those which are further apart" ([9], page 66), which Tobler paraphrased as the First Law of Geography: "Everything is related to everything else, but near things are more related than distant things" [24]. Furthermore, clusters can be associated with etiology and ecological covariates, and characteristics of a detected cluster can provide clues for isolating potential ecological factors [28]. It is therefore important to distinguish a true cluster with any shape and size from a collection of cluster candidates with regular shapes and sizes that may mask an etiological effect [3,27]. In this paper, we consider these problems in relation to the spatial scan test because it is the most frequently used test in disease surveillance [14,22,23].

Scan statistics were originally developed for one dimension point process [21] with some important asymptotic properties [4,6,10]. Kulldorff [15] extended it to two-dimensional by treating counts in spatial units as Poisson random variables. The spatial scan statistic uses a moving window of varying size to detect a set of clustered regions that are unlikely to happen by chance.

It compares disease rate within a clustered area with the rate in the rest of the area. In particular, suppose there is only one cluster in the study area, which is denoted by C_T. Then, the spatial scan statistic is to test whether C_T is a high-value or low-value cluster (i.e. a hot or a cool spot). Because C_T is unknown, it is generally to assume that C_T may belong to a collection of cluster candidates denoted by $\mathcal{C}$. A cluster candidate is often composed of a set of contiguous spatial units in the study area, which can be thought as a priori collection for spatial clusters.

As long as $\mathcal{C}$ is determined, the null hypothesis and the alternative hypothesis are

$$H_0 : \theta_C = \theta_0 \quad \text{versus} \quad H_1 : \theta_C \neq \theta_0, \tag{1.1}$$

for some $C \in \mathcal{C}$. Because the assumption that C_T belongs to $\mathcal{C}$ may be incorrect in applications, we consider the asymptotic behavior of the spatial scan statistic under the case when $C_T \in \mathcal{C}$ and $C_T \notin \mathcal{C}$, respectively. The testing problem described by (1.1) is misspecified if $C_T \notin \mathcal{C}$.

Assume a study area has m spatial units. Let Y_i be the count, y_i be the observed count, and n_i be the at-risk population in unit i, for $i = 1, \ldots, m$. Suppose Y_i are independent. Denote ϕ as the empty set, $S = \{1, \ldots, m\}$ as the set of the indices of spatial units, $\mathcal{C} = \{C_1, \ldots, C_N\}$ as the collection of cluster candidates, where $N = \#\mathcal{C}$ is the total number of cluster candidates contained in $\mathcal{C}$. Assume $C_T \neq \phi$ and $C_T \neq S$. Let θ_{0C_T} and θ_{C_T} be the expected value of Y_i for $i \in \bar{C}_T$ and $i \in C_T$, respectively, where $\bar{C}_T$ represents the complementary set of C_T. Then, the true model is

$$Y_i \sim \text{Poisson}(\theta_{0C_T} n_i), \qquad i \in \bar{C}_T; \quad \text{or} \quad Y_i \sim \text{Poisson}(\theta_{C_T} n_i), \qquad i \in C_T. \tag{1.2}$$

The difference between the hypotheses given by equation (1.1) and the true model given by equation (1.2) is that equation (1.2) is included in equation (1.1) if $C_T \in \mathcal{C}$ but not if $C_T \notin \mathcal{C}$. Our asymptotic results are derived under the true model given by equation (1.2) when the hypotheses are used as those given in equation (1.1), which include both $C_T \in \mathcal{C}$ and $C_T \notin \mathcal{C}$ cases, respectively.

Let $Y = \sum_{i=1}^m Y_i$, $y = \sum_{i=1}^m y_i$, $n = \sum_{i=1}^m n_i$, $Y_C = \sum_{i \in C} Y_i$, $y_C = \sum_{i \in C} y_i$, $n_C = \sum_{i \in C} n_i$, $Y_{\bar{C}} = \sum_{i \in \bar{C}} Y_i$, $y_{\bar{C}} = \sum_{i \in \bar{C}} y_i$, and $n_{\bar{C}} = \sum_{i \in \bar{C}} n_i$. Then, y, y_C and $y_{\bar{C}}$ are the observed values of Y, Y_C and $Y_{\bar{C}}$ under model (1.1), respectively. Under the true model, $Y_{C_T} = \sum_{i \in C_T} Y_i$ and $Y_{\bar{C}_T} = \sum_{i \in \bar{C}_T} Y_i$ are independently Poisson random variables with expected values $\theta_{C_T} n_{C_T}$ and $\theta_{0C_T} n_{\bar{C}_T}$, respectively. However, the distributions of Y_C and $Y_{\bar{C}}$ are more complicated if $C \neq C_T$.

Note that the usual alternative hypothesis of the spatial scan test assumes that

$$Y_i \sim \text{Poisson}(\theta_0 n_i), \qquad i \in \bar{C}; \quad \text{or} \quad Y_i \sim \text{Poisson}(\theta_C n_i), \qquad i \in C, \theta_C \neq \theta_0, \tag{1.3}$$

for a certain $C \in \mathcal{C}$. Since the true cluster C_T may not be contained in $\mathcal{C}$, our alternative hypothesis can be understood as a generalized version of the usual alternative hypothesis that has been considered in literature, where our alternative hypothesis is reduced to the usual alternative hypothesis if $C_T \in \mathcal{C}$. In the following, we specify the likelihood ratio statistic for a general $C \in \mathcal{C}$.

Under $H_0 \cup H_1$, the loglikelihood function is

$$\ell_C(\theta_0, \theta_C) = Y_{\bar{C}} \log \theta_0 - \theta_0 n_{\bar{C}} + Y_C \log \theta_C - \theta_C n_C + \log\left(\prod_{i=1}^m \frac{n_i^{Y_i}}{Y_i!} \right). \tag{1.4}$$

The maximum likelihood estimators (MLEs) of θ_0 and θ_C are

$$\hat{\theta}_{0C}^1 = Y_{\bar{C}}/n_{\bar{C}}$$

and

$$\hat{\theta}_C^1 = Y_C/n_C,$$

respectively, which satisfy

$$\dot{\ell}_C\big(\hat{\theta}_{0C}^1, \hat{\theta}_C^1\big) = 0,$$

where $\dot{\ell}_C(\theta_0, \theta_C)$ is the gradient of $\ell_C(\theta_0, \theta_C)$. Under H_0 for a given C, the loglikelihood function is

$$\begin{aligned} \ell_0(\theta_0) &= Y \log\theta_0 - \theta_0 n + \log\left(\prod_{i=1}^{m} \frac{n_i^{Y_i}}{Y_i!}\right), \\ \hat{\theta}_0^0 &= Y/n, \end{aligned} \tag{1.5}$$

which satisfies

$$\dot{\ell}_0\big(\hat{\theta}_0^0\big) = 0,$$

where $\dot{\ell}_0(\theta_0)$ is the derivative of $\ell_0(\theta_0)$. The logarithm of the likelihood ratio statistic is

$$\log\Lambda_C = \ell_C\big(\hat{\theta}_{0C}^1, \hat{\theta}_C^1\big) - \ell_0\big(\hat{\theta}_0^0\big).$$

The twice logarithm of the spatial scan statistic is

$$2\log\Lambda = \max_{C\in\mathcal{C}} 2\big[\ell_C\big(\hat{\theta}_{0C}^1, \hat{\theta}_C^1\big) - \ell_0\big(\hat{\theta}_0^0\big)\big], \tag{1.6}$$

where Λ is usually called the spatial scan statistic and currently a bootstrap method is often used to compute the p-value [15]. The detected cluster $\hat{C}$ then is the cluster candidate in $\mathcal{C}$ which satisfies $\Lambda_{\hat{C}} = \Lambda$. The spatial scan statistic Λ defined in (1.6) is the most frequently used test statistic to detect clusters. Currently, there is no explicit formula for $2\log\Lambda$ because it depends on the choice of $\mathcal{C}$.

As the spatial scan statistic has become increasingly popular with new extensions to capture irregular cluster shapes and ecological covariates [1,17,26,28], questions have risen regarding how to characterize a detected cluster and how to specify cluster detection among many competing conditions. For example, disease surveillance specialists often set the cluster size to be an area covering 50% of the population resulting a detected cluster that covers more than one-half of the surveillance area. To provide etiological clues, it is often necessary to repeatedly refine cluster detection size from 10% to 50% of the population size, and then reevaluate cluster strength. However, it is almost impractical to do so empirically due to a range of population size choice. In addition, it is not certain if a newly detected one from refinement would point to the same core area from the originally detected cluster. Answers to these questions require understanding theoretical properties of the spatial scan test, which is the subject of the current study.

We set out a theoretical treatment of the original and nontrivial scan test setting for which only the most important cluster is detected. According to the true distributions of Y_i specified in equation (1.2), we provide the asymptotic behaviors of the spatial scan statistic based on a set of weak regularity conditions.

- If $C_T \in \mathcal{C}$, then $\hat{C} \xrightarrow{P} C_T$; if $C_T \notin \mathcal{C}$, then $\hat{C} \xrightarrow{P} \tilde{C} \neq C_T$ for a certain $\tilde{C} \in \mathcal{C}$.
- The true distribution of $2\log\Lambda$ is asymptotically equal to the distribution of the square of the maximum of a Gaussian random field, where the mean vector of the Gaussian random field is zero under the null hypothesis and nonzero under the alternative hypothesis.
- The power function of Λ is determined by the choice of the collection of cluster candidates $\mathcal{C}$, the disease rates within and outside of C_T, and the at-risk populations within and outside of C_T.
- If the at-risk population within C_T is less than the half of the at-risk population of the whole study area, then the power function increases as the total at-risk population in the cluster increases.

In Section 2, we provide the mathematical proof of the above results as well as their rigorous statements. In Section 3, we evaluate the theoretical results via simulation and case studies. In Section 4, we provide concluding remarks.

2. Main result

We study the asymptotic properties of Λ as $n_{\min} = \min(n_1, n_2, \ldots, n_m) \to \infty$ since this reflects the case that the at risk population sizes for all units are large, which is often true in practice. We need the condition that $\lim_{n_{\min}\to\infty} n_i/n_{\min}$ positively exists because such a condition indicates that the mean (given in equation (2.10)) and covariance functions (given in equation (2.11)) in the asymptotic distribution under the local alternative uniquely exists. We assume m does not change as $n_{\min} \to \infty$ because this condition implies that the whole study area (i.e., the map) does not vary, which also indicates that both m and $\mathcal{C}$ do not vary as $n_{\min} \to \infty$. Therefore, we impose the following regularity conditions:

(a) In the true model given by (1.2), θ_{0C_T} does not vary as $n_{\min} \to \infty$, and both θ_{0C_T} and θ_{C_T} are positive.

(b) The limits $\eta_i = \lim_{n_{\min}\to\infty} \eta_{in_{\min}}$ for all $i \in S$ positively exist, where $\eta_{in_{\min}} = n_i/n_{\min}$. Therefore, $\eta_C = \sum_{i\in C} \eta_i$, $\eta_{\bar{C}} = \sum_{i\notin C} \eta_i$, and $\eta = \sum_{i=1}^{m} \eta_i$ also positively exist as $n_{\min} \to \infty$.

(c) The collection of cluster candidates $\mathcal{C}$ does not vary as $n_{\min} \to \infty$. In addition, we also have $\phi, S \notin \mathcal{C}$, and $\bar{C} \notin \mathcal{C}$ if $C \in \mathcal{C}$.

(d) Both m and N do not vary as $n_{\min} \to \infty$.

(e) Asymptotic properties are considered under one of the following conditions:

(e1) $\lim_{n_{\min}\to\infty}(\theta_{C_T} - \theta_{0C_T})$ positively exists as $n_{\min} \to \infty$; or
(e2) $\lim_{n_{\min}\to\infty} \sqrt{n_{\min}}(\theta_{C_T} - \theta_{0C_T})$ positively exists as $n_{\min} \to \infty$.

We investigated asymptotic properties of $2\log\Lambda$ under the above regularity conditions. Properties under conditions (a), (b), (c), (d), and (e1) are presented in Section 2.1. Properties under

conditions (a), (b), (c), (d), and (e2) are presented in Section 2.2. Condition (e2) is also understood as a case of local alternatives because Section 2.2 focuses on asymptotic local efficiency of the test. We use $\xrightarrow{P}$ to denote *convergence in probability* as $n_{\min} \to \infty$ and $\xrightarrow{D}$ to denote *convergence in distribution* as $n_{\min} \to \infty$, respectively.

2.1. Consistency

Let $\hat{C} \in \mathcal{C}$ be the detected cluster. Then

$$\Lambda_{\hat{C}} = \max_{C \in \mathcal{C}} \Lambda_C. \tag{2.1}$$

We say that a cluster candidate $\tilde{C}$ is *specified* asymptotically by Λ if $\hat{C} \xrightarrow{P} \tilde{C}$. We say that Λ is *consistent* if $\hat{C} \xrightarrow{P} C_T$. The consistency of the spatial scan test is whether Λ is consistent.

Denote $n_{D\cap E} = \sum_{i \in D\cap E} n_i$ and $\eta_{D\cap E} = \sum_{i \in D\cap E} \eta_i = \lim_{n\to\infty} n_{D\cap E}/n_{\min}$ for any $D, E \subseteq S$. For a deterministic sequence $\{a_{n_{\min}} : n_{\min} = 1, 2, \ldots\}$, we write $a_{n_{\min}} = o(1)$ if $\lim_{n_{\min}\to\infty} a_{n_{\min}} = 0$ or $a_{n_{\min}} = O(1)$ if $|a_{n_{\min}}|$ is uniformly bounded as $n_{\min} \to \infty$. For a stochastic sequence $\{W_{n_{\min}} : n_{\min} = 1, 2, \ldots\}$, we write $W_{n_{\min}} = o_p(1)$ if $W_{n_{\min}} \xrightarrow{P} 0$ or $W_{n_{\min}} = O_p(1)$ if $|W_{n_{\min}}|$ is uniformly bounded in probability as $n_{\min} \to \infty$. Let

$$\begin{aligned}
\theta_{0C} &= E\big(\hat{\theta}^1_{0C}\big) = \theta_{0C_T} + \frac{\eta_{C_T\cap\bar{C}}}{\eta_{\bar{C}}}(\theta_{C_T} - \theta_{0C_T}) + o(1),\\
\theta_{1C} &= E\big(\hat{\theta}^1_{C}\big) = \theta_{0C_T} + \frac{\eta_{C_T\cap C}}{\eta_{C}}(\theta_{C_T} - \theta_{0C_T}) + o(1),\\
\theta_{00} &= E\big(\hat{\theta}^0_{0}\big) = \theta_{0C_T} + \frac{\eta_{C_T}}{\eta}(\theta_{C_T} - \theta_{0C_T}) + o(1).
\end{aligned} \tag{2.2}$$

Lemma 2.1. *Assume* (a), (b), (c) *and* (d) *hold, and either* (e1) *or* (e2) *holds. If* $C \subseteq S$ *but* $C \neq \phi$ *and* $C \neq S$, *then*

$$\ell_C\big(\hat{\theta}^1_{0C}, \hat{\theta}^1_C\big) - \ell_C(\theta_{0C}, \theta_{1C}) = \frac{1}{2}\left[\frac{n_{\bar{C}}(\hat{\theta}^1_{0C} - \theta_{0C})^2}{\theta_{0C}} + \frac{n_C(\hat{\theta}^1_C - \theta_{1C})^2}{\theta_{1C}}\right] + o_p(1) = O_p(1) \tag{2.3}$$

and

$$\ell_0\big(\hat{\theta}^0_0\big) - \ell_0(\theta_{00}) = \frac{n(\hat{\theta}_0 - \theta_{00})^2}{2\theta_{00}} + o_p(1) = O_p(1). \tag{2.4}$$

Proof. Although there may be $C \neq C_T$, the results of local asymptotic normality (LAN) ([25], page 104) can always be used because we have $Y_{\bar{C}} \sim \text{Poisson}(n_{\bar{C}}\theta_{0C})$, $Y_C \sim \text{Poisson}(n_C\theta_{1C})$ and $Y \sim \text{Poisson}(n\theta_{00})$. Then, $\sqrt{n_{\bar{C}}}(\hat{\theta}^1_{0C} - \theta_{0C}) = \sqrt{n_{\bar{C}}}(Y_{\bar{C}}/n_{\bar{C}} - \theta_{0C}) \xrightarrow{D} N(0, \theta_{0C})$, $\sqrt{n_C}(\hat{\theta}^1_C - \theta_{1C}) = \sqrt{n_C}(Y_C/n_C - \theta_{1C}) \xrightarrow{D} N(0, \theta_{1C})$, and $\sqrt{n}(\hat{\theta}^0_0 - \theta_{00}) = \sqrt{n}(Y/n - \theta_{00}) \xrightarrow{D} N(0, \theta_{00})$ as

$n_{\min} \to \infty$. We compute the Taylor expansion of $\ell_C(\theta_{0C}, \theta_{1C})$ at the MLE $(\hat{\theta}^1_{0C}, \hat{\theta}^1_C)$ based on equation (1.4) and obtain

$$\begin{aligned}\ell_C(\theta_{0C}, \theta_{1C}) &= \ell_C\big(\hat{\theta}^1_{0C}, \hat{\theta}^1_C\big) - \frac{1}{2}\bigg[\frac{\bar{Y}_{\bar{C}}}{(\hat{\theta}^1_{0C})^2}\big(\hat{\theta}^1_{0C} - \theta_{0C}\big)^2 + \frac{\bar{Y}_C}{(\hat{\theta}^1_C)^2}\big(\hat{\theta}^1_C - \theta_{1C}\big)^2\bigg] + o_p(1)\\ &= \ell_C\big(\hat{\theta}^1_{0C}, \hat{\theta}^1_C\big) - \frac{1}{2}\bigg[\frac{n_{\bar{C}}}{\hat{\theta}^1_{0C}}\big(\hat{\theta}^1_{0C} - \theta_{0C}\big)^2 + \frac{n_C}{\hat{\theta}^1_C}\big(\hat{\theta}^1_C - \theta_{1C}\big)^2\bigg] + o_p(1)\\ &= \ell_C\big(\hat{\theta}^1_{0C}, \hat{\theta}^1_C\big) - \frac{1}{2}\bigg[\frac{n_{\bar{C}}}{\theta_{0C}}\big(\hat{\theta}^1_{0C} - \theta_{0C}\big)^2 + \frac{n_C}{\theta_{1C}}\big(\hat{\theta}^1_C - \theta_{1C}\big)^2\bigg] + o_p(1),\end{aligned}$$

which implies equation (2.3). We compute the Taylor expansion of $\ell_0(\theta_{00})$ at the MLE $\hat{\theta}^0_0$ based on equation (1.5) and obtain

$$\begin{aligned}\ell_0(\theta_{00}) &= \ell_0\big(\hat{\theta}^0_0\big) - \frac{Y}{2(\hat{\theta}^0_0)^2}\big(\hat{\theta}^0_0 - \theta_{00}\big)^2 + o_p(1)\\ &= \ell_0\big(\hat{\theta}^0_0\big) - \frac{n}{2\hat{\theta}^0_0}\big(\hat{\theta}^0_0 - \theta_{00}\big)^2 + o_p(1)\\ &= \ell_0\big(\hat{\theta}^0_0\big) - \frac{n}{2\theta_{00}}\big(\hat{\theta}^0_0 - \theta_{00}\big)^2 + o_p(1),\end{aligned}$$

which implies equation (2.4). □

Lemma 2.2. *Assume conditions* (a), (b), (c), (d), *and* (e1) *hold. Suppose* $C \subseteq S$ *but* $C \neq C_T$, $C \neq \phi$, *and* $C \neq S$. *Then for any* $M > 0$, *we have*

$$\lim_{n_{\min}\to\infty} P(\log \Lambda_{C_T} - \log \Lambda_C > M) = 1.$$

Proof. We can equivalently rephrase the case when $n_{\min} \to \infty$ by partitioning $Y_i = \sum_{k=1}^{n_{\min}} Y_{ik}$, where Y_{ik} are i.i.d. Poisson random variables with common mean $\theta_{0C_T}\eta_{in_{\min}}$ if $i \in C_T$ or $\theta_{1C_T}\eta_{in_{\min}}$ if $i \in C$ for all $i \in S$ and $k = 1, \ldots, n_{\min}$. Let

$$p_{C,\theta_0,\theta_C}(y_{1k}, \ldots, y_{mk}) = \bigg[\prod_{i\notin C}\frac{(\eta_{in_{\min}}\theta_0)^{y_{ik}}}{y_{ik}!}e^{-\eta_{in_{\min}}\theta_0}\bigg]\bigg[\prod_{i\in C}\frac{(\eta_{in_{\min}}\theta_C)^{y_{ik}}}{y_{ik}!}e^{-\eta_{in_{\min}}\theta_C}\bigg]$$

be the joint PMF of $(Y_{1k}, \ldots, Y_{mk})$ for a cluster candidate C, and parameters θ_0 and θ_C. The above expression is the true likelihood function if and only if $C = C_T$, $\theta_0 = \theta_{0C_T}$ and $\theta_C = \theta_{C_T}$. By the Jensen Inequality, there is

$$a_C(\theta_0, \theta_C) \le a_{C_T}(\theta_{0C_T}, \theta_{C_T}) \tag{2.5}$$

and the equality holds if and only if $C = C_T$, $\theta_0 = \theta_{0C_T}$, and $\theta_C = \theta_{C_T}$, where

$$a_C(\theta_0, \theta_C) = E\big[\ell_{C,k}(\theta_0, \theta_C)\big]. \tag{2.6}$$

For general C, θ_0, and θ_C,

$$\ell_{C,k}(\theta_0,\theta_C)=\log p_{C,\theta_0,\theta_C}(Y_{1k},\ldots,Y_{mk}) \tag{2.7}$$

is the loglikelihood function for a given k, and $\ell_{C,k}$ are i.i.d. with common expected value $a_C(\theta_0,\theta_C)$ and also with a common finite variance. The loglikelihood function for all $k=1,\ldots,n_{\min}$ is

$$\begin{aligned}
\tilde{\ell}_C(\theta_0,\theta_C)&=\sum_{k=1}^{n_{\min}}\ell_{C,k}(\theta_0,\theta_C)\\
&=\sum_{k=1}^{n_{\min}}\sum_{i\in\tilde{C}}\bigl\{Y_{ik}\bigl[\log(\eta_{in_{\min}})+\log\theta_0\bigr]-\log Y_{ik}!-\theta_0\eta_{in_{\min}}\bigr\}\\
&\quad+\sum_{k=1}^{n_{\min}}\sum_{i\in C}\bigl\{Y_{ik}\bigl[\log(\eta_{in_{\min}})+\log\theta_C\bigr]-\log Y_{ik}!-\theta_C\eta_{in_{\min}}\bigr\}\\
&=Y_{\tilde{C}}\log\theta_0-n_{\tilde{C}}\theta_0+Y_C\log\theta_C-n_C\theta_C+\sum_{i=1}^{m}Y_i\log(\eta_{in_{\min}})-\sum_{i=1}^{m}\sum_{k=1}^{n_{\min}}\log Y_{ik}!\\
&=\ell_C(\theta_0,\theta_C)-\sum_{i=1}^{m}\sum_{k=1}^{n_{\min}}\log Y_{ik}!+\sum_{i=1}^{m}\log Y_i!-Y\log n_{\min}.
\end{aligned}$$

Note that $\ell_{C,k}(\theta_0,\theta_C)$ are i.i.d. for distinct k. By the strong law of large number (SLLN) and the Shannon–Kolmogorov Information Inequality (also known as the nonegativity of the Kullback–Leibler divergence) ([8], page 113) with the limit that $\eta_i=\lim_{n_{\min}\to\infty}\eta_{in_{\min}}$ exists, we have

$$\begin{aligned}
\frac{1}{n_{\min}}\bigl[\ell_C(\theta_{0C_T},\theta_{C_T})-\ell_C(\theta_0,\theta_C)\bigr]&=\frac{1}{n_{\min}}\bigl[\tilde{\ell}_C(\theta_{0C_T},\theta_{C_T})-\tilde{\ell}_C(\theta_0,\theta_C)\bigr]\\
&=\frac{1}{n_{\min}}\sum_{k=1}^{n_{\min}}\bigl[\ell_{C_T,k}(\theta_{0C_T},\theta_{C_T})-\ell_{C,k}(\theta_0,\theta_C)\bigr] \qquad (2.8)\\
&\xrightarrow{P}a_{C_T}(\theta_{0C_T},\theta_{C_T})-a_C(\theta_0,\theta_C)\geq 0,
\end{aligned}$$

and according to (2.5) the equality holds if and only if $C_T=C$, $\theta_0=\theta_{0C_T}$, and $\theta_C=\theta_{C_T}$. Thus, we have the conclusion of the lemma. □

Theorem 2.1. *Assume conditions* (a), (b), (c), (d), *and* (e1) *hold. If* $C_T\in\mathcal{C}$, *then* Λ *is consistent. If* $C_T\notin\mathcal{C}$, *then* Λ *is inconsistent but the cluster candidate specified by* Λ *asymptotically maximizes* $a_C(\theta_{0C},\theta_{1C})$, *where* $a_C(\theta_0,\theta_C)$ *is defined by* (2.6).

Proof. Because $\mathcal{C}$ is finite, the conclusion for the case when $C_T\in\mathcal{C}$ can be directly drawn from Lemma 2.2. If $C_T\notin\mathcal{C}$, we can find $\tilde{C}$ such that $a_C(\theta_{0C},\theta_{1C})$ is maximized at $C=\tilde{C}$. If $\tilde{C}$

is unique, then $\hat{C} \xrightarrow{P} \tilde{C}$; otherwise $\hat{C}$ may not have a limit, but any convergent subsequence of $\hat{C}$ should converge to a subset of S, which is not equal to C_T but maximizes $a_C(\theta_{0C}, \theta_{1C})$. Therefore, Λ is inconsistent. In both cases, $a_C(\theta_{0C}, \theta_{1C})$ is maximized. □

Theorem 2.1 states the consistency of a spatial scan test in the two-sided alternative hypothesis. The ideas of the proof of Theorem 2.1 can be similarly used to show consistency of a spatial scan test in a one-sided alternative hypothesis, in which the alternative hypothesis in (1.1) reduces to either $H_1 : \theta_C > \theta_0$ for one hot spot only or $H_1 : \theta_C < \theta_0$ for one cool spot only. We summarize these statements into the following corollary.

Corollary 2.1. *Assume conditions* (a), (b), (c), (d), *and* (e1) *hold. If* $C_T \in \mathcal{C}$ *and* $\theta_{1C_T} > \theta_{0C_T}$, *then* Λ *is consistent under* $H_0 : \theta_C = \theta_0$ *against* $H_1 : \theta_C > \theta_0$. *If* $C_T \in \mathcal{C}$ *and* $\theta_{1C_T} < \theta_{0C_T}$, *then* Λ *is consistent under* $H_0 : \theta_C = \theta_0$ *against* $H_1 : \theta_C < \theta_0$.

2.2. Asymptotic distributions under local alternative

We use regularity condition (e2) to derive the asymptotic power function as well as the asymptotic local efficiency of $2\log\Lambda$. The asymptotic power function is determined by the relationship between C_T and $\mathcal{C}$, which is primarily decided by the total population size within the cluster. The asymptotic local efficiency is given by the behavior of the mean and covariance functions of the asymptotic distribution under the local alternative. For instance, many simulation studies (e.g., [12,13,28]) found that the probability detecting a cluster with a larger at-risk population size is higher than the probability detecting a cluster with a lower at-risk population size. Therefore, we focus on the asymptotic behavior of power functions of $2\log\Lambda$ according to the value of η_{C_T}. Our main conclusion about the asymptotic distribution is summarized in Theorem 2.2, which provides the asymptotic mean and covariance functions of $2\log\Lambda$ under the local alternative. In addition, we also provide the asymptotic local efficiency in Corollary 2.4, which concludes that the behavior of the power function of $2\log\Lambda$ is almost determined by $1/\eta_{C_T} + 1/\eta_{\tilde{C}_T}$ with minimum at $\eta_C = \eta/2$.

Let $\gamma = \sqrt{n_{\min}}(\theta_{C_T} - \theta_{0C_T})/\theta_{0C_T}$ so that $\theta_{C_T} = \theta_{0C_T}(1 + \gamma/\sqrt{n_{\min}})$. Then, γ exists and the parameters defined in (2.2) can be written into

$$\begin{aligned}
\theta_{0C} &= \theta_{0C_T}\left[1 + \left(\frac{\eta_{C_T \cap \tilde{C}}}{\eta_{\tilde{C}}}\right)\frac{\gamma}{\sqrt{n_{\min}}}\right] + o\left(\frac{1}{\sqrt{n_{\min}}}\right),\\
\theta_{1C} &= \theta_{0C_T}\left[1 + \left(\frac{\eta_{C_T \cap C}}{\eta_C}\right)\frac{\gamma}{\sqrt{n_{\min}}}\right] + o\left(\frac{1}{\sqrt{n_{\min}}}\right),\\
\theta_{00} &= \theta_{0C_T}\left[1 + \left(\frac{\eta_{C_T}}{\eta}\right)\frac{\gamma}{\sqrt{n_{\min}}}\right] + o\left(\frac{1}{\sqrt{n_{\min}}}\right).
\end{aligned} \tag{2.9}$$

Let $\mu_\gamma = (\mu_{1\gamma}, \ldots, \mu_{N\gamma})'$ be an N-dimensional vector and R be an $N \times N$ dimensional matrix, where the ith component of μ_γ is

$$\mu_{i\gamma} = \frac{\gamma\sqrt{\theta_{0C_T}}(\eta_{C_T\cap C_i}/\eta_{C_i} - \eta_{C_T\cap \bar{C}_i}/\eta_{\bar{C}_i})}{\sqrt{\sigma_{ii}}}, \tag{2.10}$$

and the (i, j)th entry of R is

$$r_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}, \tag{2.11}$$

with

$$\sigma_{ij} = \left(\frac{\eta_{C_i\cap C_j}}{\eta_{C_i}\eta_{C_j}} - \frac{\eta_{C_i\cap \bar{C}_j}}{\eta_{C_i}\eta_{\bar{C}_j}} - \frac{\eta_{\bar{C}_i\cap C_j}}{\eta_{\bar{C}_i}\eta_{C_j}} + \frac{\eta_{\bar{C}_i\cap \bar{C}_j}}{\eta_{\bar{C}_i}\eta_{\bar{C}_j}}\right).$$

Obviously, $|r_{ij}| \le 1$ and $r_{ii} = 1$ for all $i, j = 1, \ldots, m$. If conditions (a), (b), (c), (d), and (e2) hold, then $\mu_{i\gamma}$ and σ_{ij} exist as $n_{\min} \to \infty$.

Lemma 2.3. *Assume conditions* (a), (b), (c), (d), *and* (e2) *hold. Then for any* $C \subseteq S$ *but* $C \ne \phi$ *and* $C \ne S$, $2\log\Lambda_C = \delta_C^2 + o_p(1)$, *where* $\delta_C = (\hat{\theta}_C^1 - \hat{\theta}_{0C}^1)/\sqrt{\theta_{0C_T}(1/n_C + 1/n_{\bar{C}})}$.

Proof. Because $\lim_{n_{\min}\to\infty}(\theta_{C_T} - \theta_{0C_T}) = 0$, we have

$$\frac{1}{n_{\min}}\ddot{\ell}(\hat{\theta}_{0C}^1, \hat{\theta}_C^1) \xrightarrow{P} -\frac{1}{\theta_{0C_T}}\begin{pmatrix}\eta_{\bar{C}_T} & 0\\ 0 & \eta_{C_T}\end{pmatrix}$$

and

$$\frac{1}{n_{\min}}\ddot{\ell}(\hat{\theta}_0^0) \xrightarrow{P} -\frac{\eta}{\theta_{0C_T}}.$$

Straightforwardly,

$$\sqrt{n_{\min}}(\hat{\theta}_{0C}^1 - \theta_{0C_T}) \xrightarrow{D} N(0, \theta_{0C_T}/\eta_{\bar{C}_T}),$$

$$\sqrt{n_{\min}}(\hat{\theta}_C^1 - \theta_{0C_T}) \xrightarrow{D} N(0, \theta_{0C_T}/\eta_{C_T}),$$

and

$$\sqrt{n_{\min}}(\hat{\theta}_{0C}^1 - \hat{\theta}_C^1) \xrightarrow{D} N\big(0, \theta_{0C_T}(1/\eta_{\bar{C}_T} + 1/\eta_{C_T})\big).$$

Using equation (2.4) in Lemma 2.1, we have

$$\ell_0(\hat{\theta}_0^0) = \ell_0(\theta_{0C_T}) + \frac{n(\hat{\theta}_0^0 - \theta_{0C_T})^2}{2\theta_{0C_T}} + o_p(1). \tag{2.12}$$

Because $\ell_0(\theta_{0C_T}) = \ell_C(\theta_{0C_T}, \theta_{0C_T})$, using equation (2.3) in Lemma 2.1, we have

$$\ell_C\big(\hat{\theta}^1_{0C}, \hat{\theta}^1_C\big) = \ell_0(\theta_{0C_T}) + \frac{1}{2\theta_{0C_T}}\big[n_{\bar{C}}\big(\theta_{0C_T} - \hat{\theta}^1_{0C}\big)^2 + n_C\big(\theta_{0C_T} - \hat{\theta}^1_C\big)^2\big] + o_p(1).$$

Note that $n\hat{\theta}^0_0 = n_{\bar{C}}\hat{\theta}^1_{0C} + n_C\hat{\theta}^1_1$, we have

$$\begin{aligned}
2\log\Lambda_C &= 2\big[\ell_C\big(\hat{\theta}^1_{0C}, \hat{\theta}^1_C\big) - \ell_0\big(\hat{\theta}^0_0\big)\big] \\
&= \frac{1}{\theta_{0C_T}}\big[n_{\bar{C}}\big(\theta_{0C_T} - \hat{\theta}^1_{0C}\big)^2 + n_C\big(\theta_{0C_T} - \hat{\theta}^1_C\big)^2\big] - \frac{n(\theta_{0C_T} - \hat{\theta}^0_0)^2}{\theta_{0C_T}} + o_p(1) \\
&= \frac{1}{\theta_{0C_T}}\big[n_{\bar{C}}\big(\theta_{0C_T} - \hat{\theta}^1_{0C}\big)^2 + n_C\big(\theta_{0C_T} - \hat{\theta}^1_C\big)^2\big] \\
&\quad - \frac{[n_{\bar{C}}(\hat{\theta}^0_0 - \theta_{0C_T}) + n_C(\hat{\theta}^1_1 - \theta_{0C_T})]^2}{n\theta_{0C_T}} + o_p(1) \\
&= \frac{(\hat{\theta}^1_C - \hat{\theta}^1_{0C})^2}{\theta_{0C_T}(1/n_{\bar{C}} + 1/n_C)} + o_p(1),
\end{aligned}$$

which implies the conclusion of the lemma. □

Lemma 2.4. *Assume conditions* (a), (b), (c), (d), *and* (e2) *hold. Then* $\sqrt{n_{\min}}(\delta - \mu_\gamma) \xrightarrow{D} N(0, R)$, *where* $\delta = (\delta_{C_1}, \ldots, \delta_{C_N})$.

Proof. Note that N is finite. The asymptotic normality of δ can be implied by the method of the asymptotic normality of the M-estimator ([25], Theorem 5.21 on page 52). The main idea of the proof is to use M-estimation in $(\ell_{C_1}(\theta_{01}, \theta_{C_1}), \ldots, \ell_{C_N}(\theta_{0N}, \theta_{C_N}))$, where θ_{0i} are the parameters for units outside of C_i for $i = 1, \ldots, N$, respectively. The detail of proof in the verification of the conditions for the asymptotic normality of the M-estimator is straightforward but tedious. Therefore, it is omitted. Then, we only need to compute the mean and covariance function of δ.

For any $C \subseteq S$ but $C \neq \phi$ and $C \neq S$,

$$\lim_{n_{\min}\to\infty} E\big(\hat{\theta}^1_C - \hat{\theta}^1_{0C}\big) = \frac{\gamma\theta_{0C_T}}{\sqrt{n_{\min}}}\left(\frac{\eta_{C_T\cap C}}{\eta_C} - \frac{\eta_{C_T\cap\bar{C}}}{\eta_{\bar{C}}}\right).$$

For any $C, C' \subseteq S$ with $C, C' \neq \phi, S$,

$$\begin{aligned}
&\mathrm{Cov}\big(\hat{\theta}^1_C - \hat{\theta}^1_{0C}, \hat{\theta}^1_{C'} - \hat{\theta}^1_{0C'}\big) \\
&\quad = \frac{\mathrm{Cov}(Y_C, Y_{C'})}{n_C n_{C'}} - \frac{\mathrm{Cov}(Y_{\bar{C}}, Y_{C'})}{n_{\bar{C}} n_{C'}} - \frac{\mathrm{Cov}(Y_C, Y_{\bar{C}'})}{n_D n_{\bar{C}'}} + \frac{\mathrm{Cov}(Y_{\bar{C}}, Y_{\bar{C}'})}{n_{\bar{C}} n_{\bar{C}'}} \\
&\quad = \theta_{0C_T}\left(\frac{n_{C\cap C'}}{n_C n_{C'}} - \frac{n_{\bar{C}\cap C'}}{n_{\bar{C}} n_{C'}} - \frac{n_{C\cap\bar{C}'}}{n_C n_{\bar{C}'}} + \frac{n_{\bar{C}\cap\bar{C}'}}{n_{\bar{C}} n_{\bar{C}'}}\right)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{\gamma\theta_{0C_T}}{\sqrt{n_{\min}}}\left(\frac{n_{C_T\cap C\cap C'}}{n_C n_{C'}}-\frac{n_{C_T\cap\bar{C}\cap C'}}{n_{\bar{C}}n_{C'}}-\frac{n_{C_T\cap C\cap\bar{C}'}}{n_C n_{\bar{C}'}}+\frac{n_{C_T\cap\bar{C}\cap\bar{C}'}}{n_{\bar{C}}n_{\bar{C}'}}\right)\\
&\to\theta_{0C_T}\left(\frac{\eta_{C\cap C'}}{\eta_C n_{C'}}-\frac{\eta_{\bar{C}\cap C'}}{\eta_{\bar{C}}n_{C'}}-\frac{\eta_{C\cap\bar{C}'}}{\eta_C\eta_{\bar{C}'}}+\frac{\eta_{\bar{C}\cap\bar{C}'}}{\eta_{\bar{C}}\eta_{\bar{C}'}}\right)\\
&=\theta_{0C_T}\sigma_{ij}.
\end{aligned}
$$

Therefore, we have

$$
E(\delta_{C_i})=\frac{E(\hat{\theta}^1_{C_i}-\hat{\theta}^1_{0C_i})}{\sqrt{\theta_{0C_T}(1/n_{C_i}+1/n_{\bar{C}_i})}}=\frac{\gamma\sqrt{\theta_{0C_T}}(n_{C_T\cap C_i}/n_{C_i}-n_{C_T\cap\bar{C}_i}/n_{\bar{C}_i})}{\sqrt{n_{\min}(1/n_{C_i}+1/n_{\bar{C}_i})}}
$$

and

$$
\mathrm{Cov}(\delta_{C_i},\delta_{C_j})=\frac{\mathrm{Cov}(\hat{\theta}^1_{C_i}-\hat{\theta}^1_{0C_i},\hat{\theta}^1_{C_j}-\hat{\theta}^1_{0C_j})}{\theta_{0C_T}\sqrt{(1/n_{C_i}+1/n_{\bar{C}_i})(1/n_{C_j}+1/n_{\bar{C}_j})}}=\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}},
$$

which implies $\lim_{n_{\min}\to\infty}E(\delta_{C_i})=\mu_{i\gamma}$ and $\lim_{n_{\min}\to\infty}\mathrm{Cov}(\delta_{C_i},\delta_{C_j})=r_{ij}$. □

Lemma 2.5. *Let C and C' be any two nonempty subsets of $\{1,\dots,m\}$. Then,*

$$
\left|\frac{\eta_{C\cap C'}/\eta_{C'}-\eta_{C\cap\bar{C}'}/\eta_{\bar{C}'}}{\sqrt{1/\eta_{C'}+1/\eta_{\bar{C}'}}}\right|\le\frac{1}{\sqrt{1/\eta_C+1/\eta_{\bar{C}}}}
$$

and the equality holds if and only if $C=C'$ or $C=\bar{C}'$.

Proof. Straightforwardly, we have

$$
\begin{aligned}
&\left|\frac{n_{C\cap C'}}{n_{C'}}-\frac{n_{C\cap\bar{C}'}}{n_{\bar{C}'}}\right|\le\sqrt{\frac{n_C n_{\bar{C}}}{n_{C'}n_{\bar{C}'}}}\\
&\quad\Leftrightarrow\quad (n_{C\cap C'}n_{\bar{C}'}-n_{C\cap\bar{C}'}n_{C'})^2\le n_C n_{\bar{C}}n_{C'}n_{\bar{C}'}\\
&\quad\Leftrightarrow\quad (n_{C\cap C'}n_{\bar{C}\cap\bar{C}'}-n_{C\cap\bar{C}'}n_{\bar{C}\cap C'})^2\le n_C n_{\bar{C}}n_{C'}n_{\bar{C}'}.
\end{aligned}
$$

Note that $n_{C\cap C'}\le\min(n_C,n_{C'})$ and $n_{\bar{C}\cap\bar{C}'}\le\min(n_{\bar{C}},\bar{n}_{C'})$. If $n_{C\cap C'}n_{\bar{C}\cap\bar{C}'}\ge n_{C\cap\bar{C}'}n_{\bar{C}\cap C'}$, then the right-hand side of the last inequality is less than or equal to

$$
n^2_{C\cap C'}n^2_{\bar{C}\cap\bar{C}'}\le\left[\min(n_C,n_{C'})\min(n_{\bar{C}},n_{\bar{C}'})\right]^2\le n_C n_{\bar{C}}n_{C'}n_{\bar{C}'}
$$

and the equality holds if and only if $C=C'$. Note that $n_{C\cap\bar{C}'}\le\min(n_C,\bar{n}_{C'})$ and $n_{\bar{C}\cap C'}\le\min(n_{\bar{C}},n_{C'})$. If $n_{C\cap C'}n_{\bar{C}\cap\bar{C}'}\le n_{C\cap\bar{C}'}n_{\bar{C}\cap C'}$, then the right-hand side of the last inequality is less than or equal to

$$
n^2_{C\cap\bar{C}'}n^2_{\bar{C}\cap\bar{C}'}\le\left[\min(n_C,n_{\bar{C}'})\min(n_{\bar{C}},n_{C'})\right]^2\le n_C n_{\bar{C}}n'_C n_{\bar{C}'}
$$

and the equality holds if and only if $C = \bar{C}'$. We can complete the proof of the lemma by using $\eta_i = \lim_{n_{\min}\to\infty} n_i/n_{\min}$. □

Let $Z_\gamma = (Z_{1\gamma}, \ldots, Z_{N\gamma})$ be an N-dimensional normally distributed random vector with mean vector μ_γ and covariance matrix R. Denote $Z_\gamma^+ = (Z_{1\gamma}^+, \ldots, Z_{N\gamma}^+)$ and $Z^- = (Z_{1\gamma}^-, \ldots, Z_{N\gamma}^-)$, where $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$ are the positive part and negative part of x, respectively.

Theorem 2.2. *Assume conditions* (a), (b), (c), (d), *and* (e2) *hold. Then*, $2\log\Lambda$ *weakly converges to* (i) $\max_{C_i\in\mathcal{C}} Z_{i\gamma}^2$ *for any fixed* $\gamma \neq 0$ *if the alternative hypothesis is* $H_1 : \theta_C \neq \theta_0$, (ii) $\max_{C_i\in\mathcal{C}}(Z_{i\gamma}^+)^2$ *for any fixed* $\gamma > 0$ *if the alternative hypothesis is* $H_1 : \theta_C > \theta_0$, *or* (iii) $\max_{C_i\in\mathcal{C}}(Z_{i\gamma}^-)^2$ *for any fixed* $\gamma < 0$ *if the alternative hypothesis is* $H_1 : \theta_C < \theta_0$.

Proof. Let us first consider (i). Using Lemma 2.3 for a given $C_i \in \mathcal{C}$, the likelihood ratio statistic Λ_{C_i} satisfies $2\log\Lambda_{C_i} = \delta_{C_i}^2 + o_p(1)$. Because the distribution of $Z_{i\gamma}$ is the limiting distribution of δ_{C_i}, the distributions of $2\log\Lambda_{C_i}$ and $Z_{i\gamma}^2$ are asymptotically equivalent, which implies $2\log\Lambda \xrightarrow{D} \max_{C_i\in\mathcal{C}} Z_{i\gamma}^2$ because $\mathcal{C}$ is finite ([25], page 260). For (ii), the distributions of $2\log\Lambda_{C_i}$ and $(Z_{i\gamma}^+)^2$ are asymptotically equivalent, which implies $2\log\Lambda \xrightarrow{D} \max_{C_i\in\mathcal{C}}(Z_{i\gamma}^+)^2$. For (iii), the distribution of $2\log\Lambda_{C_i}$ and $(Z_{i\gamma}^-)^2$ are asymptotically equivalent, which implies $2\log\Lambda \xrightarrow{D} \max_{i\in C\in\mathcal{C}}(Z_{i\gamma}^-)^2$. □

Conclusions of Theorem 2.2 include asymptotic behaviors of power functions of $2\log\Lambda$ for $C_T \in \mathcal{C}$ and $C_T \notin \mathcal{C}$ under the local alternative, respectively. As the case when $C_T \in \mathcal{C}$ is more interesting in applications, we specify the following corollaries below.

Corollary 2.2. *Assume conditions* (a), (b), (c), *and* (d) *hold. Let* $\hat{C}$ *be the detected cluster. If* $C_T \in \mathcal{C}$, $\gamma \to \infty$, *and* $\gamma/\sqrt{n_{\min}} \to 0$ *in equation* (2.9), *then* $\hat{C} \xrightarrow{P} C_T$ *as* $n_{\min} \to \infty$.

Proof. From Lemma 2.5, for any $C \neq C_T$, we have

$$\lim_{n_{\min}\to\infty} E(Z_{C_T} - Z_C) = \gamma\sqrt{\theta_{0C_T}}\left(\frac{1}{\sqrt{(1/\eta_{C_T} + 1/\eta_{\bar{C}_T})}} - \frac{\eta_{C'\cap C}/\eta_{C'} - \eta_{\bar{C}'\cap C}/\eta_{\bar{C}'}}{\sqrt{(1/\eta_{C'} + 1/\eta_{\bar{C}'})}}\right),$$

which is positive if $\gamma > 0$ or negative if $\gamma < 0$. Therefore, $E(Z_{C_T})$ has the largest absolute value and $E^2(Z_{C_T}) - E^2(Z_C) \to \infty$ if $C \neq C_T$ as $n_{\min} \to \infty$. Because $\gamma/\sqrt{n_{\min}} \to 0$, $V(\delta_C) \to V(Z_C) = 1$ for all $C \in S$. Therefore, $\hat{C} \xrightarrow{P} C_T$. □

Corollary 2.2 implies the consistency of the spatial scan test under a weaker condition than (e1) that $\theta_{C_T} - \theta_{0C_T} = o(1/\sqrt{n_{\min}})$ as $n_{\min} \to \infty$.

Corollary 2.3. *Assume conditions* (a), (b), *and* (c) *hold. Then under the null hypothesis of* $H_0 : \theta_{C_T} = \theta_{0C_T}$, $2\log\Lambda$ *weakly converges to* (i) $\max_{C_i\in\mathcal{C}} Z_{i0}^2$ *if the alternative hypothesis is* $H_1 : \theta_C \neq \theta_0$, (ii) $\max_{C_i\in C}(Z_{i0}^+)$ *if the alternative hypothesis is* $H_1 : \theta_C > \theta_0$, *or* (iii) $\max_{C_i\in C}(Z_{i0}^-)^2$ *if the alternative hypothesis is* $H_1 : \theta_C < \theta_0$.

The asymptotic null distribution of $2\log\Lambda$ given in Corollary 2.3 depends (and only depends) on the collection of cluster candidates $\mathcal{C}$ and the ratio of at risk population sizes but not on the disease rates. Therefore, we are not able to provide a closed form formula of the asymptotic null distribution that we have stated in the corollary. To derive the asymptotic null distribution, a Monte Carlo method is used. However, as the population pattern and the collection of cluster candidates do not often change in most real world disease surveillance systems, the Monte Carlo method is only necessary to be considered once. As long as the null distribution of $2\log\Lambda$ has been derived, the Monte Carlo method is not necessary any more. Therefore, the asymptotic result given in the corollary can significantly reduce the computational burden of the test.

Corollary 2.4. *Consider the following two scenarios in the true model given by* (1.2) *when* $C_T \in \mathcal{C}$:

(i) C_{T_1} *is the cluster and* $\gamma_1 = \sqrt{n_{\min}}(\theta_{C_{T_1}} - \theta_{0C_{T_1}})/\theta_{0C_{T_1}}$ *is the true parameter for* C_{T_1};
(ii) C_{T_2} *is the cluster and* $\gamma_2 = \sqrt{n_{\min}}(\theta_{C_{T_2}} - \theta_{0C_{T_2}})/\theta_{0C_{T_2}}$ *is the true parameter for* C_{T_2}.

Assume conditions (a), (b), (c), (d), *and* (e2) *hold in the two scenarios, respectively. If*

$$\frac{\gamma_1}{\gamma_2} = \sqrt{\frac{\theta_{0C_{T_2}}(1/\eta_{C_{T_1}} + 1/\eta_{\bar{C}_{T_1}})}{\theta_{0C_{T_1}}(1/\eta_{C_{T_2}} + 1/\eta_{\bar{C}_{T_2}})}},$$

then the power functions of $2\log\Lambda$ *under* (i) *and* (ii), *respectively, are asymptotically equal.*

The proofs of Corollaries 2.3 and 2.4 are straightforward and therefore are omitted. In summary, Theorem 2.2 provides the key conclusion of the power functions under the local alternative, which can be used to numerically compute the asymptotic local efficiency. Corollary 2.3 can be used to compute the null limiting distribution and p-value of the test statistic. Because the asymptotic null distribution of $2\log\Lambda$ does not depend on θ_{0C_T}, we can use the simulation method to compute the distribution. In real world disease surveillance, the simulation method only needs to be used once because the at-risk population and the collection of cluster candidates usually do not change. Corollary 2.4 can be used to compare the behavior of power functions between two cluster specifications, which provides the asymptotic relative efficiency of the test. The theoretical properties of the spatial scan statistic that have been concluded in the theorems and corollaries are numerically evaluated via simulation and case studies in the next section.

3. Simulation and case study

In both simulation and case studies, we used real-world data. We obtained county-level lung cancer incidence data from 2003 to 2007 from the Texas Cancer Registry, and used the mid-census

population of 2005 as the at-risk population. Texas had 254 counties with a total population of 22 811 128 in 2005. The number of lung cancer incidence during the study period was 63 651, with a five-year incidents rate of 279 per 100 000.

3.1. Simulation

Under the alternative hypothesis of having one cluster only, we compared the true and asymptotic behaviors of $2\log\Lambda$, which include the specificability shown by Theorem 2.1 and the asymptotic power functions shown by Theorem 2.2. We used the county-level population distribution of Texas in 2005 as the template to design at-risk populations according to

$$n_i = \omega n_{i0}, \qquad i = 1, \ldots, 254, \tag{3.1}$$

where n_{i0} was the ith true population and n_i was the ith designed population. We introduced the designed population, which was proportional to the true population to study the consistency of $2\log\Lambda$ as $n_{\min} \to \infty$.

For a cluster candidate C, we used $n_{C0} = \sum_{i\in C} n_{i0}$ to represent the observed population within C and $n_C = \omega n_{C0}$ to represent the designed population within C. The ω value determined the variation of $n_{\min}$, and we selected ω equal to 1, 2, 5, and 10. Although $n_{\min} \to \infty$ as $\omega \to \infty$, it is sufficient to assess the behavior of the asymptotic distribution of $2\log\Lambda$ by setting $\omega = 10$, and the results for $\omega > 10$ were omitted.

We used 0.05% as the standard disease rate in our simulation because it was close to the annualized incidence rate of 55.8 per 100 000. In order to improve the speed of the simulation for comparing the proposed and the bootstrap methods, we chose $\mathcal{C}$ as the collection of adjacent counties around a cluster center. To evaluate Corollary 2.4, we designed four types of clusters in four different locations where each cluster includes the cluster center and its adjacent counties (Table 1 and Figure 1). Note that C_W was a boundary cluster. In Table 1, C_W had a small at-risk population, while C_S, C_J and C_H had larger at-risk populations.

For each designed cluster C_T (which could be either C_W, C_S, C_J, or C_H), we generated independent Poisson random samples Y_i, $i = 1, \ldots, 254$, with expected value

$$E(Y_i) = 0.0005(1 + \nu I_{i\in C_T})n_i, \tag{3.2}$$

Table 1. Designed clusters and their observed population n_{C0}

C_T	Center	Size	Counties within cluster	n_{C0}
C_W	Wichita	5	Archer, Baylor, Clay, Wichita, Wilbarger	167 512
C_S	Smith	8	Cherokee, Gregg, Henderson, Rusk Smith, Upshur, Van Zandt, Wood	607 327
C_J	Johnson	8	Bosque, Ellis, Hill, Hood, Johnson Parker, Somervell, Tarrant	2 096 477
C_H	Harris	8	Brazoria, Chambers, Fort Bend, Galveston Harris, Liberty, Montgomery, Waller	5 252 832

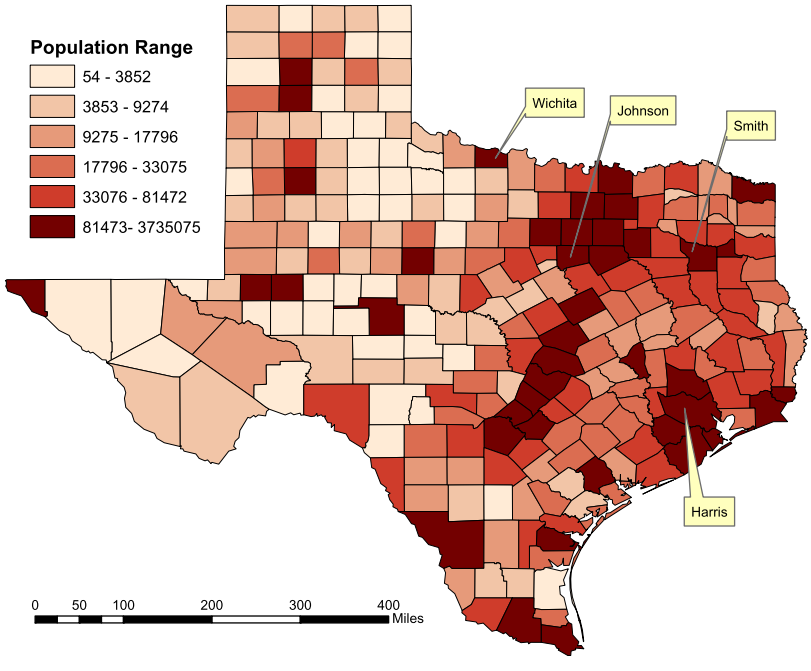

Figure 1. County-level population and designed cluster centers in Texas in 2005.

where ν increased from 0 to a certain value. For each selected ν, we generated 1000 datasets from model (3.2). We computed the null limiting distribution by using Corollary 2.3. We used the null limiting distribution and the bootstrap method to compute the p-value of $2\log\Lambda$, and to test their significance at the 0.05 level. Because power functions were approximately determined by the values of $1/\sqrt{1/n_{C_T}+1/n_{\bar{C}_T}}$ (Corollary 2.4), we used

$$\nu_a = \nu\sqrt{\frac{1/\eta_{C_W}+1/\eta_{\bar{C}_W}}{1/\eta_{C_T}+1/\eta_{\bar{C}_T}}} \tag{3.3}$$

to adjust ν in the illustration of the power functions. When we took C_T as either C_W, C_S, C_J, or C_H from Table 1, we had the values of ν_a equal to ν multiplied by 1, 1.8854, 3.3841, and 4.9310, respectively. To make the power functions between two designed clusters almost equal, a cluster with a larger at-risk population should be assigned a smaller ν value.

Location specificity is defined by the percentage of times that the detected cluster was identical to the designed cluster, and we used the percentage to assess the specificability of $2\log\Lambda$ according to Theorem 2.1. The simulation results (Table 2) supported the conclusion from Theorem 2.1. For the same cluster strength (ν), the location specificity increased as ω increased, and it would finally reach one if ω approached infinity. The location specificity was positively related to both the cluster strength and the at-risk population. In other words, a designed cluster

Table 2. Location specificity for the four designed clusters

ν	ω	C_W	C_S	C_J	C_H
0.05	1	0.013	0.059	0.185	0.279
	2	0.042	0.078	0.304	0.789
	5	0.064	0.239	0.862	0.998
	10	0.128	0.589	1.000	1.000
0.10	1	0.048	0.266	0.664	0.660
	2	0.093	0.458	0.989	1.000
	5	0.295	0.941	1.000	1.000
	10	0.738	1.000	1.000	1.000
0.15	1	0.103	0.570	0.854	0.863
	2	0.282	0.937	1.000	1.000
	5	0.763	1.000	1.000	1.000
	10	0.997	1.000	1.000	1.000
0.20	1	0.171	0.815	0.945	0.962
	2	0.561	0.999	1.000	1.000
	5	0.975	1.000	1.000	1.000
	10	1.000	1.000	1.000	1.000

with either a larger at-risk population or greater cluster strength had a higher increasing rate as ω increased. These results also supported the conclusion of Corollary 2.2: as ν increased, the location specificity of $2\log\Lambda$ also increased. The location specificity would also finally reach one as ν approached to infinity, which reenforces the conclusion of Theorem 2.1 for the consistency of $2\log\Lambda$.

Next, we compared the type I error probabilities and power functions of $2\log\Lambda$ from the limiting distribution and bootstrap methods. We computed the p-values of $2\log\Lambda$ according to the null limiting distribution from Corollary 2.3, and compared them with 1000-run bootstrap p-values. The power function was derived by the percentage of p-values less than the nominal p-value of 0.05. The type I error probabilities were obtained when $\nu = 0$.

The simulation results (Table 3) showed that the type I error probabilities were almost equal to the nominal p-values for all population patterns when $\nu = 0$. This result suggested that the p-value derived from the bootstrap method can be precisely calculated by the limiting distribution. Likewise, all the power functions behaved according to the conclusions of Theorem 2.2: the power functions increased as ν increased, and they reached 1 as ν increased to a certain value. In addition, the power increased as ω increased. For each selected ω with a fixed ν_a value, the power functions for C_W, C_S, C_J, and C_H were almost identical. This latter result is also expected from Corollary 2.4, as the power function is almost a function of ν_a, which can be interpreted as a population adjusted strength of a spatial cluster.

To briefly summarize, the behaviors of simulation for location specificity can be described by Theorem 2.1 and Corollary 2.2. The null distribution can be provided by Corollary 2.3, which can be used to compute the p-value of the test statistic. The power functions can be quantified by Theorem 2.2 and Corollary 2.4.

Table 3. Power functions for the four designed clusters, where v_a was equal to 1, 1.18854, 3.3841, and 4.9310 times v for $C_T = C_W, C_S, C_J, C_H$, respectively

		Limiting distribution				Bootstrap method			
ω	v_a	C_W	C_S	C_J	C_H	C_W	C_S	C_J	C_H
1	0.00	0.053	0.050	0.051	0.051	0.034	0.040	0.042	0.034
	0.10	0.092	0.085	0.077	0.073	0.076	0.064	0.057	0.044
	0.20	0.265	0.295	0.242	0.296	0.219	0.238	0.186	0.248
	0.30	0.690	0.701	0.692	0.740	0.631	0.632	0.628	0.675
	0.40	0.945	0.948	0.960	0.965	0.912	0.933	0.938	0.954
2	0.00	0.047	0.046	0.058	0.052	0.032	0.038	0.042	0.035
	0.10	0.136	0.127	0.135	0.136	0.106	0.099	0.088	0.110
	0.20	0.636	0.632	0.645	0.662	0.579	0.557	0.570	0.594
	0.30	0.975	0.980	0.978	0.991	0.965	0.970	0.968	0.986
	0.40	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000
5	0.00	0.049	0.051	0.057	0.054	0.047	0.041	0.035	0.034
	0.05	0.086	0.097	0.085	0.066	0.068	0.068	0.059	0.042
	0.10	0.358	0.359	0.365	0.412	0.298	0.318	0.314	0.343
	0.15	0.857	0.834	0.851	0.857	0.797	0.777	0.796	0.811
	0.20	0.991	0.991	0.990	0.996	0.984	0.987	0.979	0.995
10	0.00	0.046	0.051	0.051	0.062	0.038	0.037	0.040	0.037
	0.05	0.135	0.174	0.178	0.175	0.111	0.145	0.147	0.123
	0.10	0.771	0.780	0.800	0.814	0.714	0.734	0.752	0.757
	0.15	0.998	0.997	0.998	0.998	0.996	0.995	0.994	0.994
	0.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

3.2. Case study

We applied the elliptical spatial scan statistic for lung cancer cluster surveillance in Texas. The mathematical principles behind the spatial scan statistics are identical for circular, elliptical or any other shape, the only difference being the collection of candidate clusters $\mathcal{C}$. Following [17], we considered only a finite set of elliptical centroid coordinates, which were identical to the county centroids. For the purpose of demonstration, we considered only a few possible ellipse shapes and angles. We restricted $\mathcal{C}$ to be less than $1/2$ of the Texas counties such that our final $\mathcal{C}$ would not contain either C or $\bar{C}$ for any cluster candidate C. When there was more than one cluster, we used a stepwise procedure to detect all possible non-overlapping clusters according to [18,29].

After $\mathcal{C}$ was determined, we computed the null limiting distribution of $2\log\Lambda$ by simulations (Figure 2) as the basis for computing p-values. The limiting null distribution showed that the 5% upper quantile was 17.22, suggesting that a spatial scan statistic was significant if its value was greater than 17.22.

Overdispersion occurs when disease counts variability exceeds that of the predicted value from the Poisson model. To account for overdispersion, the Poisson assumption is often modified by

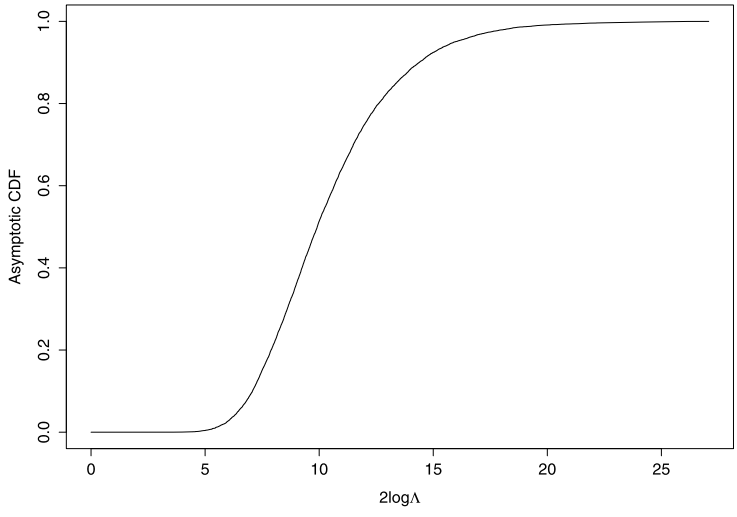

Figure 2. Cumulative distribution function of $2\log\Lambda$ for Texas cancer data when $\mathcal{C}$ is the collection of elliptical shaped cluster candidates.

the quasi-Poisson model with the marginal mean and variance relationship, in which overdispersion occurs when $V(y_i) = \xi E(y_i)$, $i = 1, \ldots, m$. Because type I error probabilities of the spatial scan tests can be significantly inflated, accounting for overdispersion is extremely helpful in the analysis and surveillance of real-world data [19]. A common way to account for overdispersion is to introduce a dispersion parameter ϕ [20], which is estimated by

$$\hat{\xi} = \max\left(1, \frac{X^2}{df_{\text{res}}}\right),$$

where X^2 is the Pearson χ^2 statistic, and df_{res} is the residual degrees of freedom. Overdispersion is present if $\hat{\xi} > 1$. When overdispersion was accounted for, we derived the modified spatial scan statistic $2\log\Lambda$ as $2\log\Lambda/\hat{\xi}$, where the definition of $2\log\Lambda$ is the same as that given in (1.6) for the pure Poisson case. When there was more than one cluster, we used a stepwise procedure to detect all possible non-overlapping clusters according to [18,29].

The result from $2\log\Lambda$ indicated the existence of several clusters. The values of $2\log\Lambda$ for the first, second, and third clusters were 3401.2, 848.7, and 386.6, respectively. All p-values from the limiting distribution and bootstrap methods were less than 0.001. After adjusting for overdispersion, however, the value of $\hat{\xi}$ was 45.07 for the first cluster, 32.01 for the second cluster, and 27.55 for the third cluster. Consequently, we had $2\log\Lambda/\hat{\xi}$ values of 75.46, 26.51, and 14.03, respectively for the first, second, and third clusters. Their corresponding p-values for the three clusters from the limiting distribution were 0, 0.0013, and 0.1831, respectively, and from the bootstrap method were 0.001, 0.002, and 0.191, respectively. Both methods suggested the existence of two clusters (Figure 3).

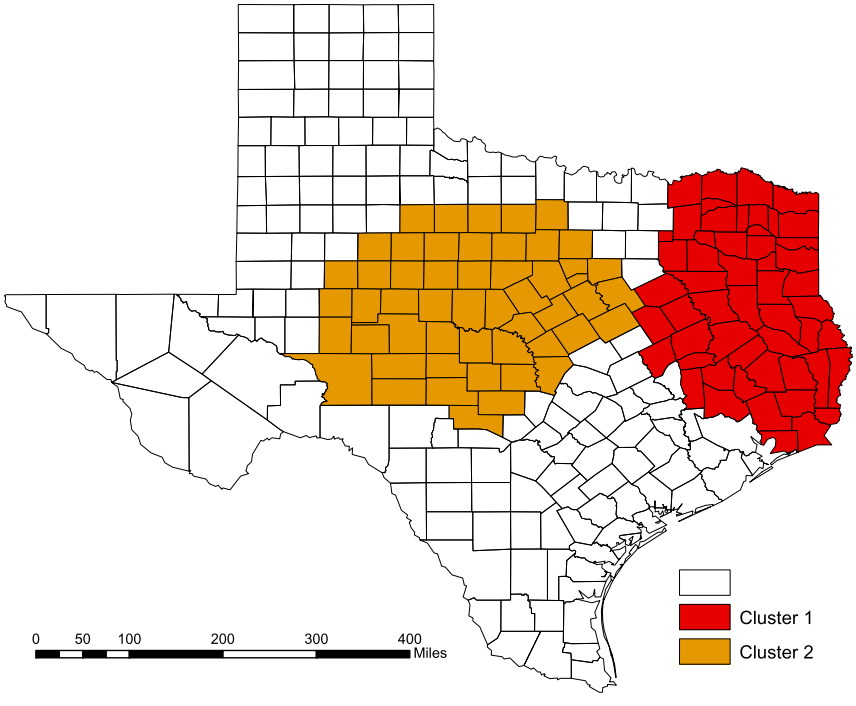

Figure 3. Spatial clusters for Texas lung cancer from 2003 to 2007.

The first cluster had 12 901 lung cancer incidents, with an at-risk population of 2 762 316. The five-year incidence rate was 467 per 100 000, which was about 67.38% higher than the state average. The center of the cluster was in the northeast part of Texas, which together with areas across the border in Oklahoma and Louisiana is known for its oil and gas industries. The second cluster had 5 777 lung cancer cases in five years with an at-risk population of 1 396 464. The five-year incidence rate was 414 per 100 000, which was 48.26% higher than the state average. This cluster was around the Hill Country, east of the Austin-Dallas corridor. Since smoking rates and COPD hospitalization rates were all high in rural part of the second cluster jackson2011, while its urban part of Dallas Metropolitan area is high on ozone concentration, this cluster may represent risks associated with these risk factors.

4. Conclusion

In this study, we have shown that the spatial scan statistic is able to detect a cluster if the cluster is included in the collection of the cluster candidates, and that the spatial scan statistic is still optimized even if the cluster is not included. The power function of the spatial scan test is asymptotically equal to the square of the maximum of a Gaussian random field, which depends

on the ratio of the at-risk population between spatial units, the collection of cluster candidates, and the strength and magnitude of the cluster. These conclusions provide a theoretical basis for future development and implementation of the spatial scan test.

Our theoretical results have several important implications. First, the finding that the limiting distribution is not unique shows a need for the computation of p-values of spatial scan tests, which will provide promising avenues to categorize limiting distributions into groups. Second, the evaluation of weak or strong clusters is likely to be resolved precisely and efficiently. Third, the asymptotic power function can provide a sound solution among clusters with different population sizes. Such complex theoretical issues are rarely touched upon in actual applications.

Given the limitation of existing methods only available in literature, we can only compare the behavior of our null limiting distribution with the bootstrap method. Even though the results from the both methods were consistent, our method can go beyond the null distribution providing a variety of cluster properties under the alternative hypothesis. Our conclusion in Theorem 2.1 can also be applied to the case when the collection of cluster candidates does not contain the true cluster, which has not been explored in literature. Our conclusion in Corollary 2.4 makes it possible to derive a population adjusted strength, which has not been used as a measure of a cluster. These conclusions can provide more information that makes further characterization of a detected cluster from impractical to practical endeavor. To conclude, our asymptotic results provides a foundation of the spatial scan test under the alternative hypothesis.

Acknowledgments

The authors appreciate suggestive comments from Associate Editor and Referees, which significantly improve the quality of the article. The research was funded by US National Science Foundation Grants SES-07-52657 (Zhang) and SES-07-52019 (Lin).

References

[1] Assunção, R., Costa, M., Tavares, A. and Ferreira, S. (2006). Fast detection of arbitrarily shaped disease clusters. *Stat. Med.* **25** 723–742. MR2225159

[2] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236. MR0373208

[3] Besag, J. and Newell, J. (1991). The detection of clusters in rare diseases. *J. Roy. Statist. Soc. Ser. A* **154** 143–155.

[4] Chan, H.P. (2009). Detection of spatial clustering with average likelihood ratio test statistics. *Ann. Statist.* **37** 3985–4010. MR2572450

[5] Commenges, D. and Jacqmin-Gadda, H. (1997). Generalized score test of homogeneity based on correlated random effects models. *J. Roy. Statist. Soc. Ser. B* **59** 157–171. MR1436561

[6] Cressie, N. (1980). The asymptotic distribution of the scan statistic under uniformity. *Ann. Probab.* **8** 828–840. MR0577319

[7] Cuzick, J. and Edwards, R. (1990). Spatial clustering for inhomogeneous populations. *J. Roy. Statist. Soc. Ser. B* **52** 73–104. MR1049303

[8] Ferguson, T.S. (1996). *A Course in Large Sample Theory*. *Texts in Statistical Science Series*. London: Chapman & Hall. MR1699953

[9] Fisher, R. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.
[10] Glaz, J., Naus, J. and Wallenstein, S. (2001). *Scan Statistics*. *Springer Series in Statistics*. New York: Springer. MR1869112
[11] Huang, L., Tiwari, R.C., Zou, Z., Kulldorff, M. and Feuer, E.J. (2009). Weighted normal spatial scan statistic for heterogeneous population data. *J. Amer. Statist. Assoc.* **104** 886–898. MR2750223
[12] Jackson, B.E., Suzuki, S., Lo, K., Su, F., Singh, K.P., Coultas, D., Bartolucci, A. and Bae, S. (2011). Geographic disparity in COPD hospitalization rates among the Texas population. *Respir. Med.* **105** 734–739.
[13] Jung, I. and Kulldorff, M. (2007). Theoretical properties of tests for spatial clustering of count data. *Canad. J. Statist.* **35** 433–446. MR2396029
[14] Kleinman, K., Abrams, A., Yih, W.K., Platt, R. and Kulldorff, M. (2006). Evaluating spatial surveillance: Detection of known outbreaks in real data. *Stat. Med.* **25** 755–769. MR2225161
[15] Kulldorff, M. (1997). A spatial scan statistic. *Comm. Statist. Theory Methods* **26** 1481–1496. MR1456844
[16] Kulldorff, M. (2006). Tests of spatial randomness adjusted for an inhomogeneity: A general framework. *J. Amer. Statist. Assoc.* **101** 1289–1305. MR2328314
[17] Kulldorff, M., Huang, L., Pickle, L. and Duczmal, L. (2006). An elliptic spatial scan statistic. *Stat. Med.* **25** 3929–3943. MR2297401
[18] Kulldorff, M., Tango, T. and Park, P.J. (2003). Power comparisons for disease clustering tests. *Comput. Statist. Data Anal.* **42** 665–684. MR1977177
[19] Loh, J.M. and Zhu, Z. (2007). Accounting for spatial correlation in the scan statistic. *Ann. Appl. Stat.* **1** 560–584. MR2415747
[20] McCullagh, P. (1983). Quasilikelihood functions. *Ann. Statist.* **11** 59–67. MR0684863
[21] Naus, J.I. (1965). The distribution of the size of the maximum cluster of points on a line. *J. Amer. Statist. Assoc.* **60** 532–538. MR0183041
[22] Odoi, A., Martin, S.W., Michel, P., Middleton, D., Holt, J. and Wilson, J. (2004). Investigation of clusters of giardiasis using GIS and a spatial statistic. *Int. J. Health Geogr.* **3** 11.
[23] Sabel, C.E., Boyle, P.J., Löytönen, M., Gatrell, A.C., Jokelainen, M., Flowerdew, R. and Maasilta, P. (2003). Spatial clustering of amyotrophic lateral sclerosis in Finland at place of birth and place of death. *Am. J. Epidemiol.* **157** 898–905.
[24] Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **46** 234–240.
[25] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247
[26] Walther, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist.* **38** 1010–1033. MR2604703
[27] Whittemore, A.S., Friend, N., Brown, B.W. Jr. and Holly, E.A. (1987). A test to detect clusters of disease. *Biometrika* **74** 631–635. MR0909368
[28] Zhang, T. and Lin, G. (2009). Spatial scan statistics in loglinear models. *Comput. Statist. Data Anal.* **53** 2851–2858. MR2667595
[29] Zhang, T., Zhang, Z. and Lin, G. (2012). Spatial scan statistics with overdispersion. *Stat. Med.* **31** 762–774. MR2901798

Received July 2014 and revised March 2015