Family of power divergence spatial scan statistics

Tonglin Zhang\textsuperscript{a,\textdagger}, Ge Lin\textsuperscript{b}

\textsuperscript{a} Department of Statistics, Purdue University, 250 North University Street, West Lafayette, IN 47907-2066, United States
\textsuperscript{b} Department of Health Services Research and Administration, College of Public Health, 984350 University of Nebraska Medical Center, Omaha, NE 68198-4330, United States

\textbf{ARTICLE INFO}

\textbf{Article history:}
Received 6 February 2013
Received in revised form 3 February 2014
Accepted 3 February 2014
Available online 7 February 2014

\textbf{Keywords:}
Cluster detection
Generalized linear models
Goodness-of-fit statistics
Power divergence family
Overdispersion
Spatial scan statistics

\textbf{ABSTRACT}

The classical spatial scan test, which derived by maximizing the likelihood ratio statistic over a collection of cluster candidates, is widely used in spatial cluster detection. As the likelihood ratio statistic is only a special case in the family of power divergence (PD) goodness-of-fit statistics, the classical spatial scan test is extended to the family of PD spatial scan tests. Therefore, the family of PD spatial scan tests includes not only the classical case but also many other cases. The test statistics, the asymptotic null distribution, and the methods to incorporate overdispersion in the cluster detection are derived. It is found that in the absence of independent variables, the asymptotic null distribution of the PD spatial scan statistics only depends on the ratio of at risks populations and the collection of cluster candidates. Particularly, the focus is on three special cases in the family. They are the deviance, the Pearson, and the freeman–Tukey spatial scan tests, where the deviance spatial scan test is equivalent to the classical spatial scan test. In simulation studies, it is found that the three test statistics are almost equally powerful for cluster detection.

© 2014 Published by Elsevier B.V.

1. Introduction

Spatial epidemiologists and spatial statisticians have long used georeferenced data for spatial pattern analysis which includes spatial cluster detection. Recent health care reform legislation in the United States mandates the use of electronic health records (EHRs) and requires that spatial data, such as patient location in latitude and longitude, be explicitly included in an EHR for meaningful use. As EHRs are implemented nationwide, more spatially referenced data will be available. A great opportunity for spatial disease investigations is to include disease surveillance, cluster detection, and etiology discovery. Kulldorff (1997)'s spatial scan statistic is widely accepted by public health agencies for disease surveillance and cluster detection. There have been many extensions of Kulldorff's spatial scan statistic that incorporate different shapes and explanatory variables, and overdispersion (Assuncao et al., 2006; Loh and Zhu, 2007; Tango and Takahashi, 2005; Zhang and Lin, 2009; Zhang et al., 2012). Most extensions are based on different statistical frameworks, resulting in parallel or even incompatible developments. In this paper, we propose an extension based on the power divergence (PD) family under the framework of generalized linear models (GLMs), which includes Kulldorff's spatial scan statistic as a special case.

Spatial data pose three difficult issues for spatial statisticians: (1) lack of repeated observations at any spatial points or spatial units, (2) high dimensionally varied and arbitrary shapes, and (3) disconnection between a variable of interest and ecological covariates. Most spatial statistical development starts addressing the first issue (Getis and Ord, 1992; Kulldorff, 1997). The original spatial scan statistic addressed the first issue by using multiple windows of varying size for cluster detection.
detection. The extensions of the spatial scan test addressed the second and third issues by accounting for multiple cluster shapes (Assuncao et al., 2006; Tango and Takahashi, 2005), ecological covariates, and overdispersion (Zhang and Lin, 2009). We propose to unify extensions to spatial scan statistic under a single framework that can serve as a platform to further develop and integrate statistics.

In this paper, we establish a methodological family that includes the spatial scan test, GLMs, and quasi-likelihood methods. We exploit the PD family (Cressie and Read, 1984; Read, 1984) and link it to a spatial scan testing method. The PD family, which is denoted by \( PD(\lambda) \), has a real index \( \lambda \). The index \( \lambda \) is considered as a selected value (Cressie and Read, 1984; Read, 1984), which includes \( PD(1) \) for the Pearson goodness-of-fit statistic, \( PD(0) \) for the deviance goodness-of-fit statistic, \( PD(-1/2) \) for Freeman–Tukey goodness-of-fit statistic. This method can include independent variables, spatial heterogeneity, and overdispersion.

Linking the spatial scan test to the PD family provides powerful options for spatial cluster detection. First, current spatial scan tests require different assessment strategies for the null hypothesis. The PD family can test a variety of hypotheses and can link the hypothesis testing and spatial scan methods together. Second, a spatial scan test, defined by a PD family, can incorporate many spatial and nonspatial effects in the cluster detection process. These effects can be accessed according to their contributions to the significance of the spatial cluster. Third, the natural connection between the PD family and GLMs provides a modeling option for cluster morphology and a platform for deriving parameter estimates and asymptotic distribution of the test statistic. In this way, any new development in spatial scan statistics can be incorporated as a member of the family.

The family of PD spatial scan statistics is proposed according to a series of GLMs, including the derivation of test statistics, estimation methods, and limiting distributions. Although the bootstrap method is often used, the limiting distribution of the test statistic provides another way to compute its \( p \)-value. In addition, methods of GLMs with mixed effects can be used to treat overdispersion and other heterogeneity effects that require model adjustments. Our simulation demonstrates the accuracy and reliability of the proposed spatial testing procedure and shows that the limiting distribution is effective and precise.

This paper is organized as follows. In Section 2, we briefly review Kulldorff’s spatial scan statistic and the PD family of goodness-of-fit statistic. After that we propose the family of PD spatial scan statistics. In Section 3, we provide asymptotic results. In Section 4, we provide numerical studies. In Section 5, we provide a discussion and a concluding remark.

2. Methods

Suppose a study area is partitioned into \( m \) spatial units. Let \( Y_i \) be the disease count, \( y_i \) be the observed count, and \( n_i \) be the at-risk population in unit \( i \), for \( i = 1, \ldots, m \). In general, \( Y_i \) is modeled by an exponential family, such as the Poisson distribution

\[
Y_i \sim \text{Poisson} (\theta_i n_i), \quad i = 1, \ldots, m, \tag{1}
\]

where \( \theta_i \) is an unknown parameter indicating a relative risk (or a disease rate), in which the observed value of \( \theta_i, y_i/n_i \), is the standard mortality rate (SMR).

2.1. Spatial scan test

Under Model (1), the spatial scan statistic searches for cluster candidates by maximizing the likelihood ratio test statistic (Kulldorff, 1997). In the simplest case, suppose that the study area has only one cluster denoted by \( C \). Let \( \bar{C} \) be the complementary set of \( C \). A statistical model for cluster detection can be specified by assuming \( \theta_i = \theta_C \) if \( i \in C \) and \( \theta_i = \theta_0 \) if \( i \in \bar{C} \). To test the existence of a cluster, one can set the null hypothesis as

\[
H_0: \theta_1 = \theta_2 = \cdots = \theta_m,
\]

where the expected count is proportional to the at-risk population (Tango, 1995; Wakefield et al., 2000). The alternative hypothesis is defined differently according to problems of interest. Overall, the alternative hypothesis may be written as:

\[
H_1: \log Y_i = \theta_0 n_i (1 + \delta_i),
\]

where \( \delta_i \) is the indicator of spatial clusters which is defined as \( \delta_i \neq 0 \) if unit \( i \) is within the cluster, \( \delta_i = 0 \) if unit \( i \) is outside of the cluster (Green and Richardson, 2002; Kulldorff, 1997). For example, if \( \delta_i = 1 \) for units within a hot spot or \( \delta_i = -1 \) for units within a cool spot, then \( H_1 \) becomes the alternative hypothesis studied by Kulldorff (1997).

Let \( \mathcal{C} \) be the collection of candidates. Denote \( N = #(\mathcal{C}) \) as the number of candidates in \( \mathcal{C} \). Then \( N \) is finite and usually much greater than \( m \). Suppose \( C \subseteq \mathcal{C} \). Denote \( Y = \sum_{i=1}^{m} Y_i, y = \sum_{i=1}^{m} y_i, n = \sum_{i=1}^{m} n_i, Y_C = \sum_{i \in C} Y_i, y_C = \sum_{i \in C} y_i, n_C = \sum_{i \in C} n_i, Y_{\bar{C}} = \sum_{i \in \bar{C}} Y_i, y_{\bar{C}} = \sum_{i \in \bar{C}} y_i, \) and \( n_{\bar{C}} = \sum_{i \in \bar{C}} n_i \), respectively. Let \( y, y_C, \) and \( y_{\bar{C}} \) be observed values of \( Y, Y_C, \) and \( Y_{\bar{C}} \), respectively. Assume \( \theta_1 = \theta_C \) if \( i \in C \) and \( \theta_i = \theta_0 \) if \( i \in \bar{C} \). Then, the likelihood ratio statistic for a test of hot spot cluster is

\[
\Lambda_C = \max_{\theta_C > \theta_0} \frac{L_C(\theta_0, \theta_C)}{\max_{\theta_C = \theta_0} L_C(\theta_0, \theta_C)} = \left( \frac{Y_C/n_C}{Y/n} \right)^{y_C} \left( \frac{Y_{\bar{C}}/n_{\bar{C}}}{Y/n} \right)^{y_{\bar{C}}}, \tag{2}
\]
when $Y_C/n_C \geq Y_{c_i}/n_C$ and $\Lambda_C = 1$ otherwise. Since $C$ is unknown, the classical spatial scan statistic is

$$\Lambda = \max_{C \in \mathcal{C}} \Lambda_C,$$

where the $p$-value of $\Lambda$ is computed by the bootstrap method.

### 2.2. Power divergence family

We focus on Poisson and binomial models in our method because they are the most often used models in the spatial scan test. As Poisson and binomial models can be interpreted by exponential families, our method is formulated under the framework of the GLM for exponential families.

Assume $Y_i$ are independently distributed of an exponential family with $\mu_i = E(Y_i)$. Let $\hat{y}_i$ be the predicted count of $y_i$. Then, the PD family goodness-of-fit statistic (Cressie and Read, 1984) is defined as

$$PD(\lambda) = \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^{m} \left( \frac{y_i}{\hat{y}_i} \right)^{\lambda} \left( \frac{\hat{y}_i}{\hat{y}_i} \right)^{\lambda - 1},$$

(4)

where $\lambda$ is a chosen number which is not considered as a parameter in the family. The cases $\lambda = 0$ and $\lambda = -1$ are defined as the limits $\lambda \to 0$ and $\lambda \to -1$, respectively. Assume a testing problem is considered with the null hypothesis being denoted by $H_0$ and the alternative hypothesis being denoted by $H_1$. Let $\hat{y}_{i,0}$ and $\hat{y}_{i,1}$ be the $i$th predicted counts under the null hypothesis $H_0$ and the alternative hypothesis $H_1$, respectively. Let $PD_0(\lambda)$ and $PD_1(\lambda)$ be the values of $PD(\lambda)$ under $H_0$ and $H_0 \cup H_1$, respectively. When $\hat{y}_i$ in Eq. (4) is replaced by $\hat{y}_{i,0}$ and $\hat{y}_{i,1}$, respectively, the difference between $PD_0(\lambda)$ and $PD_1(\lambda)$ is

$$PD_d(\lambda) = PD_0(\lambda) - PD_1(\lambda) = \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^{m} \left( \frac{\hat{y}_{i,0}^{\lambda+1}}{\hat{y}_{i,0}^{\lambda+1}} - \frac{\hat{y}_{i,1}^{\lambda+1}}{\hat{y}_{i,1}^{\lambda+1}} \right).$$

(5)

Under $H_0$ for a given $C$, $PD_d(\lambda)$ approximately follows a $\chi^2$-square distribution. The null hypothesis is rejected if $PD_d(\lambda)$ is large.

The PD family includes many well-known goodness-of-fit statistics (Cressie and Read, 1984), including the Pearson goodness-of-fit statistic

$$X^2 = PD(1) = \sum_{i=1}^{m} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i},$$

(6)

the deviance goodness-of-fit statistic

$$G^2 = PD(0) = \sum_{i=1}^{m} 2y_i \log(y_i/\hat{y}_i),$$

(7)

and the Freeman–Tukey goodness-of-fit statistic

$$F^2 = PD \left( -\frac{1}{2} \right) = \sum_{i=1}^{m} 4(\sqrt{y_i} - \sqrt{\hat{y}_i})^2.$$  

(8)

In real applications, any of $PD(\lambda)$ can be used to test $H_0$ against $H_1$ in the problem, where the $p$-value of $PD(\lambda)$ is derived from the asymptotic $\chi^2$ distribution under the null hypothesis.

### 2.3. Family of PD spatial scan statistics

Assume a GLM for count is specified as:

$$g(\mu_i) = \xi_i + X_i' \beta + \alpha_i, \quad i = 1, \ldots, m,$$

(9)

where $g(\cdot)$ is a known link function, $\xi_i$ is an offset term, $X_i = (x_{i1}, \ldots, x_{ip})$ with $x_{i1} = 1$ is a vector of independent variables, $\beta = (\beta_1, \ldots, \beta_p)'$ is a vector of unknown parameters, $\alpha_i$ is an indicator of a cluster defined as $\alpha_i = \alpha_C$ if $i \in C$ and $\alpha_i = 0$ if $i \notin C$. It can be seen that Model (1) is a special case of Model (9).

Under the null hypothesis of $H_0 : \alpha_C = 0$, Model (9) becomes

$$g(\mu_i) = \xi_i + X_i' \beta.$$  

(10)

Denote $\mu_{ic} = \mu_i$ for a cluster $C$. Under the alternative hypothesis, Model (9) becomes

$$g(\mu_{ic}) = \xi_i + X_i' \beta + \alpha_C I_{i \in C}.$$  

(11)
where $\alpha_C$ indicates the cluster strength and direction. If $\alpha_C > 0$, then C is a hot spot. If $\alpha_C < 0$, then C is a cool spot. In Eq. (11), we assume that $\beta$ is the same for all units in the study area. However, the estimates of $\beta$ in Eqs. (10) and (11) are usually different. Therefore, we use $\hat{\beta}$ to denote the estimator of $\beta$ in Eq. (10) and $\hat{\beta}_C$ to denote the estimator of $\beta$ in (11).

Let $\hat{\beta}$ be the MLE of $\beta$ under Model (10), and $\hat{\alpha}_C$ and $\hat{\beta}_C$ be the conditional MLEs of $\alpha_C$ and $\beta$ under Model (11) for a given $C$. Then, $\hat{\alpha}_C$ and $\hat{\beta}_C$ as well as the predicted counts $\hat{y}_C$ under $H_0$ and $\hat{y}_{iC}$ under $H_1$ can all be derived easily (Agresti, 2002; Faraway, 2006). Let $PD_0(\lambda)$ and $PD_{IC}(\lambda)$ be the values of $PD(\lambda)$ under Model (10) and Model (11), respectively. The $PD(\lambda)$ test statistic is

$$PD_C(\lambda) = PD_0(\lambda) - PD_{IC}(\lambda).$$  \hspace{1cm} (12)

Then, the $PD(\lambda)$ family spatial scan statistic is

$$PD_C(\lambda) = \max_{C} PD_C(\lambda).$$  \hspace{1cm} (13)

The null hypothesis is rejected if the value of $PD_C(\lambda)$ is large.

**Example 1.** Kulldorff’s spatial scan statistic under Poisson distribution. Let us start with the original spatial scan statistic that tests only one hot spot without any independent variable. In this case, we have $\xi = \log(n_i)$, $p = 1$, and $\alpha_C > 0$ in Model (11), with $g$ is the log link function. The MLE of $\theta_C$ under the null hypothesis is $\hat{\theta}_0 = y/n$. The MLEs of $\theta_0$ and $\theta_C$ under the alternative hypothesis are $\hat{\theta}_0C = y_C/n_C$ and $\hat{\theta}_C = y_C/n_C$ if $y_C/n_C \geq y/n$, or $\hat{\theta}_0C = \hat{\theta}_0 = y/n$ otherwise. Therefore, $\hat{\gamma}_C = \hat{\theta}_0n_i, \hat{\gamma}_{iC} = \hat{\theta}_Cn_i$ if $i \in C$, and $\hat{\gamma}_{iC} = \hat{\theta}_0Cn_i$ if $i \notin C$.

$$PD_C(0) = PD_0(0) - PD_{IC}(0) = 2 \left[ y_C \log \left( \frac{y_C/n_C}{y/n} \right) + y_C \log \left( \frac{y_C/n_C}{y/n} \right) \right],$$

if $y_C/n_C \geq y_C/n_C$ and $PD_C(0) = 0$, otherwise. In this case, $PD_C(0) = 2 \log \Lambda_C$, where $\Lambda_C$ is given by Eq. (2), which implies that $PD_C(0) = 2 \log \Lambda$. Therefore, $PD_C(0)$ is equivalent to Kulldorff’s spatial scan statistic.

**Example 2.** The Pearson and Freeman–Tukey spatial scan statistics under Poisson distribution. Under the same assumptions of Example 1, when $\lambda = 1$,

$$PD_C(1) = PD_0(1) - PD_{IC}(1) = \sum_{i=1}^{m} \frac{(y_i - n_i \xi/n)^2}{n_i \xi/n} - \sum_{i \in C} \frac{(y_i - n_i \xi_C/n_C)^2}{n_i \xi_C/n_C} - \sum_{i \notin C} \frac{(y_i - n_i \xi_C/n_C)^2}{n_i \xi_C/n_C},$$

if $y_C/n_C \geq y_C/n_C$ and $PD_C(1) = 0$ otherwise. This is the Pearson $\chi^2$ spatial scan statistic. When $\lambda = -1/2$,

$$PD_C \left( -\frac{1}{2} \right) = PD_0 \left( -\frac{1}{2} \right) - PD_{IC} \left( -\frac{1}{2} \right) = 2 \left[ \sum_{i \in C} \sqrt{\gamma_i n_i} (\sqrt{\gamma_i/n} - \sqrt{y_C/n_C}) + \sum_{i \notin C} \sqrt{\gamma_i n_i} (\sqrt{\gamma_i/n} - \sqrt{y_C/n_C}) \right],$$

if $y_C/n_C \geq y_C/n_C$ and $PD_C(-1/2) = 0$ otherwise. This is the Freeman–Tukey spatial scan statistic.

**Example 3.** Poisson distribution with overdispersion. Overdispersion can be incorporated in the quasi-Poisson model (McCullagh, 1983) with the mean and variance in the form of

$$V(Y_i) = \phi E(Y_i), \quad i = 1, \ldots, m,$$  \hspace{1cm} (14)

where $\phi$ is the dispersion parameter. Overdispersion is not present if $\phi = 1$. When overdispersion is present ($\phi > 1$), the dispersion parameter $\phi$ should be estimated. For example, we may use a classical method which estimates $\phi$ by

$$\hat{\phi} = \max \left( \frac{X^2}{df_{res}}, 1 \right),$$  \hspace{1cm} (15)

where $X^2 = \sum_{i=1}^{m} (y_i - \hat{\gamma}_0)^2/\hat{\gamma}_0$ is the Pearson $\chi^2$ statistic and $df_{res}$ is the residual degree of freedom. Particularly, we can modify Eq. (12) by

$$PD_{C,\phi}(\lambda) = \frac{PD_0(\lambda) - PD_{IC}(\lambda)}{\hat{\phi}}, \quad C \in \mathbb{C},$$  \hspace{1cm} (16)
and Eq. (13) by
\[ PD_s(\lambda) = \max_{C \subseteq c} PD_{C,s}(\lambda). \]  

(17)

The PD_{s,o}(\lambda) under Model (14) can account for overdispersion and reduce type I error probabilities (Zhang and Lin, 2009; Zhang et al., 2012). This, together with the flexibility of including independent variables, can reduce the effect of other potential spatial heterogeneity.

**Example 4.** Binomial distribution. If \( Y_i \sim Bin(n_i, \theta_i) \), we may choose the logistic, the probit, or the complementary log–log link as in Model (10). Then,
\[ PD_s(\lambda) = \max_{C \subseteq c} \left\{ \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^{m} \left[ \frac{y_i^{\hat{y}+1}}{\hat{y}_i} - \frac{y_i^{\hat{y}+1}}{\hat{y}_i} + \frac{(n_i - y_i)^{\hat{y}+1}}{(n_i - \hat{y}_i)^2} - \frac{(n_i - y_i)^{\hat{y}+1}}{(n_i - \hat{y}_i)^2} \right] \right\}. \]

We can, similarly, derive the likelihood ratio scan statistic PD_s(0), the Pearson scan statistic PD_s(1), and the Freeman–Tukey scan statistic PD_s(−1/2). Likewise, we can derive the PD scan test for negative binomial distribution when \( Y_i \sim NB(\mu_i, k) \) with \( E(Y_i) = \mu_i \) and \( V(Y_i) = \mu_i + \mu_i^2/k \) for a given k.

**Example 5.** Poisson distribution for a secondary cluster with an independent variable. Consider the Poisson data with the log link in Model (11). Suppose an independent variable \( x_i \) is included, and the first cluster \( C_1 \) has been identified. We choose \( p = 3 \) with \( \xi_i = \log(n_i), x_{13} = 1 \) if \( i \in C_1 \) and \( x_{13} = 0 \) if \( i \in C_1 \), Model (10) becomes
\[ \log(\theta_i) = \beta_1 + \beta_2 x_i + \beta_3 l_{iC_1}, \]
and Model (11) becomes
\[ \log(\theta_i) = \beta_1 + \beta_2 x_i + \beta_3 l_{iC_1} + \alpha_C l_{iC}, \]
where \( C_1 \) is known but \( C \) is not. Then, PD_s(\lambda) can be used to identify the second cluster \( C \) with independent variable \( x_i \).

### 3. Asymptotic theory

The asymptotic distribution of PD_s(\lambda) is considered under the null hypothesis of no spatial cluster as \( n_{min} = \min(n_1, n_2, \ldots, n_m) \to \infty \) mainly for the Poisson or binomial distribution with \( E(Y_i) = \theta_i n_i \). Let \( \theta_0 = (\theta_{01}, \ldots, \theta_{0p})' \) be the true value of \( \beta \) in Model (10). Then, \( \alpha_C = 0 \) under Model (11). If \( Y_i \) follows a Poisson distribution with the log link and \( \xi_i = \log(n_i) \), then the true parameter is
\[ \theta_0 = e^{\xi_0}. \]

If \( Y_i \) follows a binomial distribution with the logistic link and \( \xi_i = 1 \), then the true parameter is
\[ \theta_0 = e^{\xi_0} / (1 + e^{\xi_0}). \]

The property of PD_s(\lambda) is derived under the pure Poisson or binomial model and the Poisson or binomial model with overdispersion, respectively. In the pure Poisson model, \( Y_i \) exactly follows a Poisson distribution with \( E(Y_i) = V(Y_i) = n_i \theta_i \). In the pure binomial model, \( Y_i \) exactly follows a binomial distribution with \( E(Y_i) = n_i \theta_i \) and \( V(Y_i) = n_i \theta_i (1 - \theta_i) \). Overdispersion is not present in either the pure Poisson or binomial model. When overdispersion is present, the true distribution of \( Y_i \) is not exactly Poisson or binomial any more. To describe overdispersion for the Poisson model, the classical method is to introduce a dispersion parameter \( \phi \) such that
\[ V(Y_i) = \phi E(Y_i). \]

In other words, Eq. (20) may be \( V(Y_i) = \phi n_i \theta_0 \). In the binomial model, a similar model may be \( V(Y_i) = \phi n_i \theta_0 (1 - \theta_0) \). In the following of this paper, we first derive the limiting distribution of PD_s(\lambda) in the pure Poisson and binomial model and then extend the result to the case when overdispersion is present, which are given in Theorem 1 and Corollary 1, respectively.

We first consider the pure Poisson and binomial models. We assume \( \alpha_C \) and \( \beta \) are not restricted so that \( \alpha_C \in \mathbb{R} \) and \( \beta \in \mathbb{R}^p \) in Model (11). We denote \( \hat{\beta} \) as the MLE of \( \beta \) in Model (10) and (\( \hat{\alpha}_C, \hat{\beta}_C \)) as the MLE of \( (\alpha_C, \beta) \) in Model (11). We assume that the intercept term is always contained in \( x \). Then, when the maximum likelihood method is used to estimate model parameters in (9), we always have \( \sum_{i=1}^{m} \hat{y}_i = \sum_{i=1}^{m} y_i \) in Eq. (4) (Agresti, 2002, p. 138). We impose the following regularity conditions, which are common in most statistical models.

**Regularity conditions for the pure Poisson or binomial model.**

(C0) The random count \( Y_i \) is either independently Poisson distributed with log link or binomial distributed with logistic, probit or complementary link;
(C1) The true value $\beta_0$ is a vector of finite real numbers in Model (10);
(C2) For any $C \in C$, the Jacobian matrix

$$j(\alpha_C, \beta) = \begin{pmatrix}
\frac{\partial \ell^2_0}{\partial \alpha_C} & \frac{\partial \ell^2_0}{\partial \alpha_C \partial \beta_1} & \cdots & \frac{\partial \ell^2_0}{\partial \alpha_C \partial \beta_p} \\
\frac{\partial \ell^2_0}{\partial \beta_1} & \frac{\partial \ell^2_0}{\partial \beta_1 \partial \alpha_C} & \cdots & \frac{\partial \ell^2_0}{\partial \beta_1 \partial \beta_p} \\
\cdots & \cdots & \cdots & \cdots \\
\frac{\partial \ell^2_0}{\partial \beta_p} & \frac{\partial \ell^2_0}{\partial \beta_p \partial \alpha_C} & \cdots & \frac{\partial \ell^2_0}{\partial \beta_p \partial \beta_p}
\end{pmatrix}$$

has a full rank of $p + 1$ at $\alpha_C = 0$;
(C3) The ratios $r_i = n_i/n_{\min}$ for all $i = 1, \ldots, m$ do not change as $n_{\min} \to \infty$;
(C4) The true parameter $\beta_0$ and the independent vector $x$, do not change as $n_{\min} \to \infty$;
(C5) Both $\mathcal{C}$ and $\tilde{\mathcal{C}}$ are not empty for $C \in \mathcal{C}$.

Let

$$A_{Cj_Ck} = \frac{\sum_{i \in C_j \cap C_k} r_i \theta_{i0}}{\left( \sum_{i \in C_j} r_i \theta_{i0} \right) \left( \sum_{i \in C_k} r_i \theta_{i0} \right)}.$$  \hspace{1cm} (21)

where $\theta_{i0}$ is defined by Eq. (18) for the Poisson data or by Eq. (19) for the binomial data, and $C_j \in \mathcal{C}$ or $\tilde{C}_j \in \mathcal{C}$ and $C_k \in \mathcal{C}$ or $\tilde{C}_k \in \mathcal{C}$. Let $Z = (Z_1, Z_2, \ldots, Z_N)$ be an $N$-dimensional normally distributed random vector with zero mean and variance–covariance matrix $V$, where the $(j, k)$-th entry of $V$ is given by

$$v_{jk} = \frac{A_{Cj_Ck} - A_{Cj_{\tilde{C}k}} - A_{Cj_{C\tilde{k}}} + A_{Cj_{\tilde{C}\tilde{k}}}}{(A_{Cj_Cj} + A_{Cj_{\tilde{C}}} A_{Cj_{C\tilde{k}}} A_{Cj_{\tilde{C}\tilde{k}}})^{1/2}}.$$  \hspace{1cm} (22)

Then, $v_{jj} = 1$ and $|v_{jk}| < 1$ if $j \neq k$. We write $Z^+_j = \max(Z_j, 0)$ and $Z^-_j = \max(-Z_j, 0)$ as the positive part and the negative part of $Z_j$, respectively. We state the theorems below and provide the proofs in the Appendix.

**Theorem 1.** Assume (C1)–(C5) hold. Let $n_{\min} \to \infty$. Then, $PD_s(\lambda) \xrightarrow{D} \max_{C \in \mathcal{C}} Z_j^2$. Therefore, the $p$-value of the test can be approximated by $P[PD_s(\lambda) \geq x] = \max_{C \in \mathcal{C}} Z_j^2$ if the alternative hypothesis includes both hot and cool spots; $P[PD_s(\lambda) \geq x] = P[\max_{C \in \mathcal{C}} (Z_j^+) \geq x]$ if the alternative hypothesis includes hot spots only; or $P[PD_s(\lambda) \geq x] = P[\max_{C \in \mathcal{C}} (Z_j^-) \geq x]$ if the alternative hypothesis includes cool spots only.

**Theorem 1** provides the limiting distribution of $PD_s(\lambda)$ for the pure Poisson and binomial case. It cannot be used to compute the limiting distribution when overdispersion is not present. When overdispersion is present, we need to consider the PD-family spatial scan statistic with the adjustment of dispersion parameter, which is

$$PD_{s, o}(\lambda) = \max_{C \in \mathcal{C}} PD_{s, o}(\lambda) = \frac{1}{\phi} \max_{C \in \mathcal{C}} PD_{s}(\lambda) = \frac{1}{\phi} PD_{s}(\lambda). \hspace{1cm} (23)$$

Eq. (23) is an extension of Eq. (16) in the pure Poisson or binomial case.

**Corollary 1.** Assume all conditions of **Theorem 1** hold. Suppose there is an estimator $\hat{\phi}$ of dispersion parameter $\phi$ such that $\hat{\phi} \xrightarrow{p} \phi$ as $n_{\min} \to \infty$. Then, $PD_{s, o}(\lambda) \xrightarrow{D} \max_{C \in \mathcal{C}} Z_j^2$.

**Proof.** The result can be directly implied by quasi-Poisson method introduced by McCullagh (1983) with Slutsky’s Theorem (Ferguson, 1996, p. 39). □

Denote

$$A_{Cj_Ck} = \frac{\sum_{i \in C_j \cap C_k} r_i}{\sum_{i \in C_j} \sum_{i \in C_k} r_i} \sum_{i \in C_j} r_i.$$
for either $C_j \in \mathcal{C}$ or $\tilde{C}_j \in \mathcal{C}$ and either $C_k \in \mathcal{C}$ or $\tilde{C}_k \in \mathcal{C}$. Let $\tilde{Z} = (\tilde{Z}_1, \tilde{Z}_2, \ldots, \tilde{Z}_N)$ be normal random vectors with mean 0 and variance–covariance matrix $\tilde{V}$, where the $(j, k)$-th entry of $\tilde{V}$ is

$$
\tilde{v}_{jk} = \frac{\tilde{A}_{C_j C_k} - \tilde{A}_{C_j \tilde{C}_k} - \tilde{A}_{\tilde{C}_j C_k} + \tilde{A}_{\tilde{C}_j \tilde{C}_k}}{(\tilde{A}_{C_j C_j} + \tilde{A}_{C_j \tilde{C}_j} + \tilde{A}_{\tilde{C}_j C_k} + \tilde{A}_{\tilde{C}_j \tilde{C}_k})^{1/2}}.
$$

Then, $\tilde{v}_{jj} = 1$ and $|\tilde{v}_{jk}| < 1$ for $j \neq k$.

**Corollary 2.** Assume (C1)–(C5) hold. Suppose there is no independent variable in Models (10) and (11), respectively. Then $\Lambda \overset{D}{\to} \max_{\theta \in \mathcal{C}} \tilde{Z}_i^2$.

Both Theorem 1 and Corollary 2 can be used to compute the asymptotic null distribution of $PD_1(\lambda)$. Because the idea in Corollary 2 is more easily to be stated than the idea in Theorem 1, we focus on the interpretation of our idea in Corollary 2 below.

When independent variables are not present, the asymptotic null distribution of $PD_1(\lambda)$ given in Corollary 2 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 Corollary 2. In order 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 $PD_1(\lambda)$ has been derived, the Monte Carlo method is not necessary any more. Therefore, the asymptotic result given in Corollary 2 can significantly reduce the computational burden in a disease surveillance system. In the following, we propose a Monte Carlo method to compute the asymptotic null distribution of $PD_1(\lambda)$. The method to the randomization procedure to generate Monte Carlo replication of the null and alternative hypothesis was defined originally in Dwass (1957). The condition $n_{\min} \to \infty$ that we have considered in our asymptotic result is represented by $\theta_0 \to \infty$ in the algorithm, where $\theta_0$ is the average disease rate in the whole study area.

Monte Carlo method for the asymptotic null distribution of $PD_1(\lambda)$ when independent variables are not present:

(i) Select a $\theta_0$ and assume disease rates in the spatial units are all equal to $\theta_0$. Then, choose a large value of $\eta$ in $n'_i = \eta n_i$.
(ii) Independently generate $Y_i \sim \text{Poisson}(\theta_0 n'_i)$ for $i = 1, \ldots, m$.
(iii) Compute the value of $PD_1(\lambda)$ for a selected $\lambda \in \mathbb{R}$.
(iv) Repeat steps (ii) and (iii) many times. The asymptotic null distribution of $PD_1(\lambda)$ can be derived by using the sample distribution of the simulated values of $PD_1(\lambda)$.

The value of $\theta_0$ is not necessary equal to its observed value in the algorithm because the asymptotic null distribution is derived as $n_{\min} \to \infty$, which is represented by $\eta \to \infty$. The asymptotic null distribution of $PD_1(\lambda)$ is independent of $\lambda$, which can be used to compute the p-value of $PD_1(\lambda)$. The asymptotic null distribution can be used if neither the ratio of at risk population nor the collection of cluster candidates has significant changes.

There are a few limitations when the asymptotic null distribution is used to compute the p-value. First, even though we have derived the asymptotic distribution of $PD_1(\lambda)$, we have not provided a closed form formula in the method. Using the Monte Carlo method to compute the asymptotic null distribution may reduce the speed of the method. Therefore, it is necessary to provide an efficient method to compute the asymptotic null distribution. Second, ratios of at risk population between units and the collection of cluster candidates are contained in the formula of the asymptotic null distribution. If they have significant changes, then the asymptotic null distribution of $PD_1(\lambda)$ should be re-computed. Third, the asymptotic null distribution is derived under the condition that $n_{\min} \to \infty$. Using the asymptotic null distribution in the computation of p-value may cause a bias if the number of cases is low, especially when there are many zero cases in spatial units. Therefore, it is important to examine the influence of low number of cases in the method.

4. Numerical studies

The purpose of our numerical studies was to compare the performance of $PD_1(\lambda)$ for different values of $\lambda$. As $PD_1(0)$ is the classical spatial scan statistic, $PD_1(\lambda)$ for any $\lambda \neq 0$ provides an alternative method of the classical spatial scan test. As the Pearson, deviance, and Freeman–Tukey statistics are the most often used goodness-of-fit statistics, we focused on $PD_1(\lambda)$ for $\lambda = 1, 0, -1/2$, which corresponds to the Pearson, deviance, and Freeman–Tukey spatial scan statistics, respectively.

4.1. Simulation

We evaluated the performance of our $PD_1(\lambda)$ by simulating several population and cluster patterns on the $10 \times 10$ lattice. Therefore, we had $m = 100$ spatial units. We considered the following four at-risk population patterns ranging
from homogeneous (a) to heterogeneous (b)–(d), with different cluster locations:
(a) we set \( n_i = 10^4 \) for all \( i \), and then inserted a cluster \( C \) centered at \(( 5, 5 )\) with radius \( \sqrt{2} \);
(b) we set \( n_i = 10^5 \) if the distance between the \( i \)-th point and the point \((7, 7)\) was less than or equal to 2 and \( n_i = 10^4 \) otherwise, and then inserted a cluster \( C \) centered at the point \((3, 3)\) with radius \( \sqrt{2} \);
(c) we used the same population pattern in (b), but moved the center of cluster \( C \) to the point \((5, 5)\);
(d) we used the same population pattern in (b), but moved the center of cluster \( C \) to the point \((7, 7)\);

In above settings, (a) had a homogeneous population pattern; (b)–(d) had the same heterogeneous population patterns but their cluster patterns were different. In particular, (b) had the cluster in a sparsely populated (e.g., rural) area; (c) had the cluster at the boundary of highly populated (e.g., suburban) area; and (d) had the cluster inside of a highly populated (e.g., downtown) area.

By setting \( \lambda = 1, 0, -1/2, \) PD\(_i\)(\( \lambda \)) corresponds to the Pearson \( X_i^2 \), deviance \( G_i^2 \), and Freeman–Tukey \( F_i^2 \) spatial scan statistics, respectively. Since \( G_i^2 \) is equivalent to Kulldorff’s spatial scan statistic, we chose Kulldorff’s method as the baseline and compared the Type I error probabilities and power functions for different values of \( \lambda \).

To compare type I error probabilities and power functions of the three PD\(_i\)(\( \lambda \)) statistics, we generated the random counts \( Y_i \) independently from the Poisson distribution with expected value equal to

\[
E(Y_i|\omega_i) = 0.001(1 + \delta I_{\text{I} \text{I} \text{I} \text{L}})n_i\omega_i, \quad i = 1, \ldots, 100.
\]

where \( \delta \) was the measure of the strength of the spatial cluster and \( \omega_i \) with \( E(\omega_i) = 1 \) was a random effect for a measurement of overdispersion. When \( \delta = 0 \), spatial cluster effect was not present. When \( V(\omega_i) = 0 \) such that \( \omega_i = 1 \) for all \( i \), overdispersion was not present. When \( \delta \) was larger, the model had a stronger spatial cluster effect. To decide the magnitude of overdispersion, we imposed a parameter \( \phi \) according to the quasi-Poisson model described by McCullagh (1983). In this model, we generated \( \omega_i \) independently from \( T(\alpha, b) \) with \( a_i = b_i = 0.001(1 + \delta I_{\text{I} \text{I} \text{I} \text{L}})n_i/\phi - 1 \) such that \( E(\omega_i) = 1 \) and \( V(\omega_i) = (\phi - 1)/[0.001(1 + \delta I_{\text{I} \text{I} \text{I} \text{L}})n_i] \). After \( \omega_i \) had been generated, generated \( Y_i \) independently from Poisson distribution with conditional expected value given in Eq. (24). Therefore, we had

\[
E(Y_i) = E(Y_i|\omega_i) = 0.001(1 + \delta I_{\text{I} \text{I} \text{I} \text{L}})n_iE(\omega_i)
\]

and

\[
V(Y_i) = E[V(Y_i|\omega_i)] + V[E(Y_i|\omega_i)]
= 0.001(1 + \delta I_{\text{I} \text{I} \text{I} \text{L}})n_iE(\omega_i) + [0.001(1 + \delta I_{\text{I} \text{I} \text{I} \text{L}})n_i]^2V(\omega_i)
= 0.001(1 + \delta I_{\text{I} \text{I} \text{I} \text{L}})n_i\phi
= \phi E(Y_i).
\]

In our simulation, the overdispersion was present if \( \phi > 1 \) and the true rate was 0.001 if clusters were not present.

Before comparing the three test statistics, we examined the true null distribution against the asymptotic null distribution of PD\(_i\)(\( \lambda \)) under the null hypothesis. To derive the true null distribution, we generated \( 10^4 \) random samples from the Poisson distribution with expected value 0.001 \( n_i \) (rate was 0.001). We computed the values of PD\(_i\)(\( \lambda \)) for \( \lambda = 1, 0, -1/2 \). We used the sample distribution of these \( 10^4 \) simulated values of PD\(_i\)(\( \lambda \)) as the true null distribution of PD\(_i\)(\( \lambda \)). To derive the asymptotic null distribution, we generated \( 10^4 \) random samples from the Poisson distribution with expected value equal to 0.1 \( n_i \) (rate was 0.1 which was equivalent to \( \eta = 100 \)) in the algorithm given at the end of Section 3. We compared the curves of the two sample distributions in Fig. 1, where the upper panel is based on the population pattern in (a) and the lower panel is based on the population pattern in (b). The three limiting distributions of \( X_i^2 \), \( G_i^2 \), and \( F_i^2 \) were clearly aligned with the distributions from the true distribution for both population samples. The curves for \( G_i^2 \) and \( F_i^2 \) were close to but the curves for \( X_i^2 \) were a little bit away from the straight line, with 45° across the origin indicating that the limiting distribution is trustworthy for \( G_i^2 \) and \( F_i^2 \). To understand the curves, we computed the 95% upper quantiles of the true distributions and the asymptotic null distributions for \( X_i^2 \), \( G_i^2 \), and \( F_i^2 \) and found that they were all less than 12.20. Therefore, we concluded that it was precise to use the asymptotic null distribution in the computation of p-values for \( G_i^2 \) and \( F_i^2 \), and it would not cause a significant bias in the computation for \( X_i^2 \).

We used \( X_i^2 \), \( G_i^2 \), and \( F_i^2 \) to detect the spatial clusters according to the four spatial designs above. We used \( k \)-nearest neighbors (\( k = 2 \) to 50) to detect the clusters, such that the size of a cluster was restricted to greater than one but less than half of the simulation area. We generated 1000 simulated data from Model (24) for \( \delta \). We computed the rejection rates of the selected statistics based on 1000 simulation repetitions for \( \delta \) varied from 0 to 0.5. For each repetition, we used both the limiting distribution and the bootstrap method to derive the p-value of PD\(_i\)(\( \lambda \)). We claimed the cluster was significant if the p-value was less than 0.05. We found that the results of the limiting distribution and the bootstrap methods were almost identical.

The results in Table 1 were derived under Model (24) when \( \delta = 0 \). It showed that the type I error probabilities of the three statistics were acceptable and close to \( \alpha = 0.05 \) for all four patterns in the absence of overdispersion (\( \phi = 1 \)). However, when overdispersion was present (\( \phi = 2 \)), the rejection rates of PD\(_i\)(\( \lambda \)) were inflated but those of PD\(_{0.0}(\lambda) \) were not. Therefore, the adjustment of overdispersion was efficient.
Fig. 1. Comparison of the quantile functions between the true distribution and limiting and true distributions of $PD_s(\lambda)$, where both quantile functions were derived by a simulation with $10^4$ replications. The dashed line represents the straight line with slope 1 and intercept 0.

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>Pattern</th>
<th>$X_s^2$</th>
<th>$C_s^2$</th>
<th>$F_s^2$</th>
<th>$X_{s,0}^2$</th>
<th>$C_{s,0}^2$</th>
<th>$F_{s,0}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(a)</td>
<td>0.047</td>
<td>0.045</td>
<td>0.043</td>
<td>0.037</td>
<td>0.042</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>0.048</td>
<td>0.048</td>
<td>0.042</td>
<td>0.040</td>
<td>0.045</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>(c)</td>
<td>0.049</td>
<td>0.051</td>
<td>0.047</td>
<td>0.039</td>
<td>0.041</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>(d)</td>
<td>0.050</td>
<td>0.055</td>
<td>0.059</td>
<td>0.043</td>
<td>0.051</td>
<td>0.049</td>
</tr>
<tr>
<td>2</td>
<td>(a)</td>
<td>0.565</td>
<td>0.519</td>
<td>0.534</td>
<td>0.048</td>
<td>0.056</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>0.532</td>
<td>0.495</td>
<td>0.509</td>
<td>0.041</td>
<td>0.047</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>(c)</td>
<td>0.539</td>
<td>0.521</td>
<td>0.534</td>
<td>0.048</td>
<td>0.049</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>(d)</td>
<td>0.550</td>
<td>0.524</td>
<td>0.517</td>
<td>0.044</td>
<td>0.054</td>
<td>0.053</td>
</tr>
</tbody>
</table>

Table 1
Type I error probabilities of $PD_s(\lambda)$ and $PD_{s,0}(\lambda)$.

We also compared the power functions of $X_s^2$, $C_s^2$, and $F_s^2$ by taking $\delta > 0$ in Model (24) without overdispersion, which showed that the power functions of $F_s^2$ were slightly higher than the power functions of $X_s^2$ or $C_s^2$ for moderate or large clusters (Fig. 2). The power functions approached 0.05 as $\delta$ went to 0 because they became the type I error probabilities when $\delta = 0$, and increased to 1 as $\delta$ became large. The power functions of the local cluster at highly populated areas were higher than those at sparsely populated areas. This result was expected because a cluster at highly populated areas is more reliable and less sensitive to small changes in incidence.

4.2. Application

We assessed the proposed method in the context of disease surveillance for breast, colorectal, and lung cancers in Texas. The three cancer sites contribute the most to Texas cancer mortality (Zhan and Lin, 2003). We obtained cases counts for
2003–2007 from the Texas Cancer Registry, and used the mid-census population of 2005 as the at-risk population. We derived the testing results from the asymptotic distribution, and the bootstrap methods and the results were close.

In 2005, Texas had 254 counties with a total population of 22,811,128. The five-year incidence was 109 per 100,000 for breast cancer, 50.8 for lung cancer, and 40 for colorectal cancer. Both incidence and mortality of individual cancer sites have been studied (Hsu et al., 2006). Excessive lung cancer mortality and a higher than average incidence of colorectal cancer were found in northeastern Texas. We wanted to investigate whether the cluster locations, as separately reported by others, were similar. Exploratory analysis suggested there were substantial overdispersion among incidents for the three cancer sites: 7.88, 18.07, and 44.90 for breast, colorectal, and lung cancers, respectively. We therefore used both $PD_s(\lambda)$ and $PD_{s,o}(\lambda)$ methods, where $PD_s(\lambda)$ and $PD_{s,o}(\lambda)$ were computed by using (13) and (17), respectively.

We computed the asymptotic null distribution (Fig. 3) of $PD_s(\lambda)$ from Corollary 2 with the algorithm given at the end of Section 3, where the disease rate was chosen as $\theta_0 = 0.1$ in the Monte Carlo method with $10^6$ replications. We computed the $p$-values of $PD_s(\lambda)$ and $PD_{s,o}(\lambda)$ from the asymptotic null distribution. The $p$-values were compared with the $p$-values.
Fig. 4. Texas cancer clusters and county-level population distribution.

Table 2
Breast (B), lung (L) and colorectal (C) cancer clusters for Texas counties based on $G^2_s$ and $G^2_{s,o}$, where the $p$-values of $G^2_s$ were all less than 0.001.

<table>
<thead>
<tr>
<th>Center</th>
<th>Size</th>
<th>$(y_c, n_c)$</th>
<th>$\hat{\delta}$</th>
<th>$G^2_s$</th>
<th>$G^2_{s,o}$ ($p$-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>1st</td>
<td>Shelby</td>
<td>43 (7688, 1193333)</td>
<td>0.1846</td>
<td>234.5</td>
</tr>
<tr>
<td></td>
<td>2nd</td>
<td>Shackelford</td>
<td>46 (3280, 486498)</td>
<td>0.2396</td>
<td>147.85</td>
</tr>
<tr>
<td>L</td>
<td>1st</td>
<td>Shelby</td>
<td>50 (12306, 2618340)</td>
<td>0.6844</td>
<td>3289.79</td>
</tr>
<tr>
<td></td>
<td>2nd</td>
<td>Nolan</td>
<td>90 (7657, 1945520)</td>
<td>0.4105</td>
<td>896.61</td>
</tr>
<tr>
<td>C</td>
<td>1st</td>
<td>Panola</td>
<td>48 (6132, 2073707)</td>
<td>0.4717</td>
<td>902.20</td>
</tr>
<tr>
<td></td>
<td>2nd</td>
<td>TomGreen</td>
<td>79 (3040, 1092680)</td>
<td>0.3487</td>
<td>364.05</td>
</tr>
</tbody>
</table>

from the bootstrap method. We found that the results were almost the same. Therefore, we decided to only show the results from the bootstrap method because of its popularity.

The results showed that all the three cancer sites had a primary cluster in northeastern Texas, with a very significant $p$-value, and the cluster sizes differed by 7 or 10 counties at most (Fig. 4). Even though the first set of clusters was significant for all cancer sites, lung cancer had the strongest effect, with a $G^2_{s,o}$ of 73.28, while breast cancer was the weakest, with a $G^2_{s,o}$ of 29.76 (Table 2). However, all clusters were located in northeastern Texas, where the state's population of African Americans is highly concentrated, and where the economy is dominated by oil and gas industries. In addition, a secondary cluster was found in central or north central Texas, but significance levels for the three cancer sites were all close to 0.05, and cluster sizes varied greatly, from 42 for breast cancer, to 90 for lung cancer. It should be pointed out that the results from $F^2_{s,o}$ and $X^2_{s,o}$ were also consistent with those from $G^2_{s,o}$.

We display the $p$-values of the bootstrap method in Table 2 due to that method's popularity, but using the asymptotic distribution would have almost the same result. The advantages of using asymptotic null distribution are (1) to efficiently compare different cluster detection strategies, (2) to explore theoretical properties of cluster detection, and (3) to investigate the behavior of cluster detection within and across spatial patterns. With the aid of asymptotic distribution, simulation studies and perspective analysis can be designed and organized explicitly.
5. Discussion

The PD family provides a tool set of various test statistics, and the GLM bridges the tool set and other spatial statistical methods. Our contribution is the derivation of the PD family spatial scan tests, their asymptotic properties, and the connection to the GLM. The GLM not only bridges our proposed method, but also provides linkage to other model-based methods, such as ecological covariates and dispersion parameters, which are commonly not available in spatial statistical tests. Reciprocally, sophisticated or elaborate spatial search methods often enable spatial statistical methods to generate spatial indicator variables, and they can be incorporated naturally in GLMs to uncover unknown relationships. Furthermore, our contribution complements model-based geostatistics (Diggle et al., 1998), which has already included spatial random field and other variables. We used a spatial scan test as an example, but the idea can be extended to many other spatial statistical methods.

GLMs provide a general framework to include various factors in cluster detection. One of most common concerns in disease surveillance practices is inflated false alarm rates. By incorporating a dispersion parameter, we effectively reduced type I error probabilities from over 50% to around 5%. Note that the overdispersion effect used in the evaluation was modest (φ = 2), and such an effect is common in real-world data. By effectively removing some overdispersion effect, the spatial scan statistic should find more robust applications. In addition, our method can potentially incorporate risk factors and different shape parameters in cluster detection by treating them as independent variables.

Our approach to linking a PD family via GLMs to spatial scan tests opens a wide field for future work. Although we applied only three members of a PD family, any member can be applied to a spatial cluster test, and some may have greater benefits than others. Identifying the best member for a particular spatial cluster test is valuable. The linkage of spatial cluster detection methods to GLMs also opens the door for future work to improve cluster detection efficiency. Our previous research suggests accounting for overdispersion will greatly increase location specificity and cluster detection accuracy (Zhang and Lin, 2009).

Acknowledgments

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

Appendix. Proofs

Let \( \hat{y}_i \) and \( \hat{y}_{iC} \) be the \( i \)th predicted count under Models (10) and (11) respectively. Let \( PD_0(\lambda) \) and \( PD_{1C}(\lambda) \) as the values of \( PD(\lambda) \) under Models (10) and (11), respectively. Then \( PD_{1C}(\lambda) \) is \( PD(\lambda) \) when \( y_i \) is replaced by \( \hat{y}_{iC} \), and \( PD_0(\lambda) \) is \( PD(\lambda) \) when \( y_i \) is replaced by \( \hat{y}_i \). We describe the detail of the proofs for Poisson distribution with log link and simply mention the proofs for binomial distribution with logistic link. The proofs for binomial distribution with probit or complementary links are similar. Thus, we omit them.

A.1. Proofs for Poisson distribution with log link

If \( Y_i \) are independent Poisson random variables and \( g(t) = \log(t) \) is the log link, then the log-likelihood function under Model (10) is

\[
\ell(\beta) = - \sum_{i=1}^{m} Y_i! + \sum_{i=1}^{m} Y_i(\log n_i + \mathbf{x}_i')\beta) - \sum_{i=1}^{m} n_i e^{\mathbf{x}_i'\beta},
\]

and the log-likelihood function under Model (11) is

\[
\ell_C(\alpha, \beta) = - \sum_{i=1}^{m} Y_i! + \sum_{i=1}^{m} Y_i \left( \log n_i + \sum_{j=1}^{p} x_{ij}\beta_j + \alpha c_{iC} \right) - \sum_{i=1}^{m} n_i e^{\sum_{j=1}^{p} x_{ij}\beta_j + \alpha c_{iC}}.
\]

Then \( \hat{\beta} \) and \( (\hat{\alpha}_C, \hat{\beta}_C) \) are derived by maximizing \( \ell(\beta) \) and \( \ell_C(\alpha, \beta) \), respectively.

**Lemma 1.** Under regularity conditions (C1)-(C5) , \( \hat{\alpha}_C \overset{P}{\to} 0, \hat{\beta}_C \overset{P}{\to} \beta_0, \hat{\beta} \overset{P}{\to} \beta_0, \) and as \( n_{\min} \to \infty \) their limiting distributions are

\[
\sqrt{n_{\min}} \left[ \begin{array}{c} \hat{\alpha}_C \\ \hat{\beta}_C \\ \hat{\beta} \end{array} \right] \overset{D}{\to} N(0, I^{-1}),
\]

where \( I_C \) is the \((p + 1) \times (p + 1)\) Fisher Information matrix, and

\[
\sqrt{n_{\min}}(\hat{\beta} - \beta_0) \overset{D}{\to} N(0, I^{-1}),
\]

(A.1)
where $I$ is the $p \times p$ Fisher Information matrix. The $(i, j)$-th entry of $I$, denoted by $I_{ij}$, for $k, l = 0, 1, \ldots, p$, is $I_{ij} = \sum_{l=0}^{p} r_l \theta_l \theta_l$, $I_{ij} = \sum_{i=0}^{m} r_i x_i \theta_i \theta_i$, or $I_{i0} = \sum_{i=0}^{m} r_i x_i \theta_i$. The $(i, j)$-th entry of $I$, denoted by $I_{ij}$ for $k = 1, \ldots, n$, is $I_{ij} = I_{ij}$.

**Proof.** It is sufficient to show the consistency and asymptotic normality of $(\hat{\alpha}_C, \hat{\beta}_C)$. Those of $\hat{\beta}$ can be shown similarly.

In order to derive the consistency and the asymptotic normality of the MLE, we equivalently rephrase the case when $n_{\min} \to \infty$ by assuming $Y_l$ are independently Poisson distributed with expected value $E(Y_l) = n_{l0} e^{\sum_{j=1}^{p} x_j \beta_j + \alpha c_l x_l}$.

for $i = 1, \ldots, m$ and $l = 1, \ldots, L$. We assume $n_{l0}$ are fixed and $n_l = L n_{l0}$. Suppose the random vector $\tilde{Y}_l = (Y_{l1}, Y_{l2}, \ldots, Y_{ln})$ are iid and $Y_l = \sum_{i=1}^{n} Y_{li}$. Then, $n_{\min} \to \infty$ can be understood as $L \to \infty$.

For a given $l$, the likelihood function based on $\tilde{Y}_l$ is

$$\tilde{L} = \tilde{L}(\alpha_C, \beta_C; Y_{l1}, \ldots, Y_{ln})$$

$$= - \sum_{i=1}^{n} \log(Y_{li}) + \sum_{i=1}^{n} \log(n_{l0}) + \sum_{i=1}^{n} Y_{li} \left( \sum_{j=1}^{p} x_j \beta_j + \alpha C_l x_l \right) - \sum_{i=1}^{n} n_{l0} e^{\sum_{j=1}^{p} x_j \beta_j + \alpha c_l x_l}.$$ 

Then, the log-likelihood function based on $Y_l$ for all $i = 1, \ldots, m$ and $l = 1, \ldots, L$ is

$$\tilde{L}(\alpha_C, \beta_C) = \frac{1}{n} \sum_{i=1}^{n} \log(n_{l0}) + \sum_{i=1}^{n} Y_{li} \left( \sum_{j=1}^{p} x_j \beta_j + \alpha C_l x_l \right) - \sum_{i=1}^{n} n_{l0} e^{\sum_{j=1}^{p} x_j \beta_j + \alpha c_l x_l}$$

The MLE $(\hat{\alpha}_C, \hat{\beta}_C)$ can be equivalently solved by maximizing $\tilde{L}(\alpha_C, \beta_C)$. Consistency and asymptotic normality of $(\hat{\alpha}_C, \hat{\beta}_C)$ can be derived by using the method of iid samples (Ferguson, 1996, p. 112). To compute the asymptotic variance of $(\hat{\alpha}_C, \hat{\beta}_C)$, we need to calculate the Fisher Information matrix based on $\tilde{L}(\alpha_C, \beta_C)$. The result is below.

The first order partial derivatives of $\tilde{L}(\alpha_C, \beta_C)$ are

$$\frac{\partial \tilde{L}}{\partial \alpha_C} = \frac{1}{n} \sum_{i=1}^{n} Y_{li} - e^{\alpha C_l x_l} \sum_{i=1}^{n} n_{l0} e^{\sum_{j=1}^{p} x_j \beta_j},$$

and

$$\frac{\partial \tilde{L}}{\partial \beta_k} = \frac{1}{n} \sum_{i=1}^{n} Y_{li} x_i - \sum_{i=1}^{n} n_{l0} e^{\sum_{j=1}^{p} x_j \beta_j + \alpha c_l x_l} x_i, \quad k = 1, \ldots, p.$$ 

The second order partial derivatives of $\tilde{L}(\alpha_C, \beta_C)$ are

$$\frac{\partial^2 \tilde{L}}{\partial \alpha_C^2} = - e^{\alpha C_l x_l} \sum_{i=1}^{n} n_{l0} e^{\sum_{j=1}^{p} x_j \beta_j},$$

$$\frac{\partial^2 \tilde{L}}{\partial \alpha_C \partial \beta_k} = - e^{\alpha C_l x_l} \sum_{i=1}^{n} n_{l0} e^{\sum_{j=1}^{p} x_j \beta_j} x_i, \quad k = 1, \ldots, p,$$

and

$$\frac{\partial^2 \tilde{L}}{\partial \beta_k \partial \beta_{k'}} = - \sum_{i=1}^{n} n_{l0} e^{\sum_{j=1}^{p} x_j \beta_j + \alpha c_l x_l} x_i x_i, \quad k_1, k_2 = 1, \ldots, p.$$ 

As $L \to \infty$, we have

$$\sqrt{L} \left[ \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right] \xrightarrow{D} N(0, I^{-1}(0, \theta_0)).$$
where $\tilde{I}(0, \beta_0)$ is the $(p + 1) \times (p + 1)$ Fisher Information matrix. Its entries are

$$\tilde{I}_{00} = \sum_{i \in C} n_0 e^{-\sum_{j=1}^p x_{ij} \beta_0},$$

$$\tilde{I}_{0k} = \tilde{I}_{k0} = \sum_{i \in C} n_0 e^{-\sum_{j=1}^p x_{ij} \beta_0} x_{ik}, \quad k = 1, \ldots, p,$$

and

$$\tilde{I}_{k1,k_2} = \sum_{i = 1}^m n_0 e^{-\sum_{j=1}^p x_{ij} \beta_0} x_{ik} x_{ik_2}, \quad k_1, k_2 = 1, \ldots, p.$$

The limiting distribution of $(\hat{\alpha}_C, \hat{\beta}_C)$ given by Eq. (A.1) can be derived if $L = n_{\text{min}}$. The consistency and the asymptotic normality of $\beta$ can be shown similarly. □

**Lemma 2.** Assume (C1)–(CS) hold. For a $C \in C$ or $\tilde{C} \in C$, let $R_1 = \sum_{i \in C} r_i e^{\hat{x}_i \hat{\beta}_C} / \sum_{i \in C} r_i \beta_0$, $R_2 = \sum_{i \in \tilde{C}} r_i e^{\hat{x}_i \hat{\beta}_C} / \sum_{i \in \tilde{C}} r_i \beta_0$ and $R_3 = R_1 - R_2$. Then, $R_1 \xrightarrow{p} 1$, $R_2 \xrightarrow{p} 1$ and $R_3 \xrightarrow{p} 0$ as $n_{\text{min}} \to \infty$.

**Proof.** This can be derived from Lemma 1 by using the Continuous Mapping Theorem. □

**Lemma 3.** Assume (C1)–(CS) hold. For a fixed $\lambda$ and for a given $C \in C$,

$$PD_C(\lambda) = \frac{(\hat{\alpha}_C - \alpha_0)}{\sigma_{\hat{\alpha}_C}}^2 + o_p(1),$$

as $n_{\text{min}} \to \infty$, where $o_p$ is the standard deviation of $\hat{\alpha}_C$ given by Eq. (A.1).

**Proof.** Because $PD_0(0)$ is the likelihood ratio goodness-of-fit statistic, the conclusion holds when $\lambda = 0$ (van der Vaart, 1998, p. 231). Note that $\sum_{i=1}^m \hat{y}_i = \sum_{i=1}^m \hat{y}_{iC} = \sum_{i=1}^m \hat{y}_i$. When $\lambda = -1$, by Taylor expansion, we have

$$PD_{-1} = -2 \sum_{i=1}^m \hat{y}_i \log \left( \frac{y_i}{\hat{y}_i} \right)$$

$$= -2 \sum_{i=1}^m \hat{y}_i \log \left( 1 + \frac{y_i - \hat{y}_i}{\hat{y}_i} \right)$$

$$= -2 \sum_{i=1}^m \hat{y}_i \left\{ \frac{y_i - \hat{y}_i}{\hat{y}_i} - \frac{1}{2} \left( \frac{y_i - \hat{y}_i}{\hat{y}_i} \right)^2 + O_p \left( \frac{(y_i - \hat{y}_i)^3}{\hat{y}_i} \right) \right\}$$

$$= \sum_{i=1}^m \left( \frac{y_i - \hat{y}_i}{\hat{y}_i} \right)^2 + o_p(n_{\text{min}}^{-1/2})$$

Similarly, we have $PD_{0}(-1) = PD_{0}(1) + o_p(1)$. Therefore, we have

$$PD_{-1} = PD_{0}(-1) - PD_{-1} = PD_{0}(1) + o_p(1).$$

When $\lambda \neq 0, -1$, by Taylor expansion, we have

$$PD_{\lambda}(\lambda) = \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^m y_i \left( \frac{y_i}{\hat{y}_i} \right)^{\lambda} - 1$$

$$= \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^m \hat{y}_i \left( \frac{y_i}{\hat{y}_i} \right)^{\lambda + 1} - 1$$

$$= \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^m \hat{y}_i \left( 1 - \frac{y_i - \hat{y}_i}{\hat{y}_i} \right)^{\lambda + 1} - 1$$

$$= \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^m \hat{y}_i \left\{ (\lambda + 1) \left( \frac{y_i - \hat{y}_i}{\hat{y}_i} \right) + \frac{\lambda(\lambda + 1)}{2} \left( \frac{y_i - \hat{y}_i}{\hat{y}_i} \right)^2 + O_p \left( \frac{(y_i - \hat{y}_i)^3}{\hat{y}_i} \right) \right\}$$

$$= \sum_{i=1}^m \hat{y}_i \left( \frac{y_i - \hat{y}_i}{\hat{y}_i} \right)^2 + \frac{\sigma_{\hat{\alpha}_C}}{\sigma_{\hat{\alpha}_C}^2} \sum_{i=1}^m \hat{y}_i \left( \frac{y_i - \hat{y}_i}{\hat{y}_i} \right)^2 + O_p(n_{\text{min}}^{-1/2})$$

$$= PD_{\lambda}(1) + o_p(1).$$
Similarly, we have $PD_0(\lambda) = PD_0(1) + o_p(1)$. Therefore,

$$PD_{C}(\lambda) = PD_0(\lambda) - PD_{1C}(\lambda) = PD_0(1) + o_p(1), \quad \lambda \neq 0, -1.$$ 

Combine those, we have $PD_{C}(\lambda) = PD_0(1) + o_p(1)$ for all $\lambda \neq 0$. Because $PD_{C}(1) = PD_0(1) + o_p(1)$ (van der Vaart, 1998, p. 245), we have

$$PD_{C}(\lambda) = \left(\frac{\hat{\alpha}_C}{\sigma_{\alpha C}}\right)^2 + o_p(1),$$

for all $\lambda \in R$. □

**Lemma 4.** Assume (C1)–(C5) hold. Let $\alpha = (\alpha_{C_1}, \alpha_{C_2}, \ldots, \alpha_{C_l})$ for distinct $C_1, C_2, \ldots, C_l \in \mathcal{C}$ and $\hat{\alpha} = (\hat{\alpha}_{C_1}, \hat{\alpha}_{C_2}, \ldots, \hat{\alpha}_{C_l})$ be its MLE. Then

$$\sqrt{n_{\min}}\hat{\alpha} \overset{D}{\rightarrow} N(0, \Sigma),$$

where $\Sigma$ is a $l \times l$ matrix with $(j, k)$ entry given by

$$\sigma_{jk} = A_{C_j C_k} - A_{C_j \tilde{C}_k} - A_{\tilde{C}_j C_k} + A_{\tilde{C}_j \tilde{C}_k},$$

for $j, k = 1, 2, \ldots, l$.

**Proof.** Note that $Y_c = \sum_{i \in C} Y_i \sim \text{Poisson}(n_i e^{x_i' \theta_0})$. Let $U_C = (Y_c - \sum_{i \in C} n_i \theta_0) / \sum_{i \in C} n_i \theta_0$ for $C \in \mathcal{C}$ or $\tilde{C} \in \mathcal{C}$. Then, $U_C$ and $U_{\tilde{C}}$ are independent, and

$$\sqrt{n_{\min}}U_C \overset{D}{\rightarrow} N(0, A_{CC}),$$

as $n_{\min} \rightarrow \infty$. Note that $x_{ii} = 1$ for all $i = 1, \ldots, m$. Then, for a given $C$, we have

$$\frac{\partial \ell_C(\alpha, \beta)}{\partial \beta_{1C}} = Y - e^\alpha \sum_{i \in C} n_i e^{x_i' \beta} - \sum_{i \in C} n_i e^{x_i' \beta},$$

and

$$\frac{\partial \ell_C(\alpha, \beta)}{\partial \alpha_C} = Y_c - e^\alpha \sum_{i \in C} n_i e^{x_i' \beta}.$$ 

Combining the two equations and using **Lemma 2**, we have

$$e^{\hat{\alpha}_C} - 1 = \frac{Y_c}{\sum_{i \in C} n_i e^{x_i' \beta}} - \frac{Y_{\tilde{C}}}{\sum_{i \in \tilde{C}} n_i e^{x_i' \beta}}$$

$$= R_1 \left( \frac{Y_C}{\sum_{i \in C} n_i \theta_0} \right) - R_2 \left( \frac{Y_{\tilde{C}}}{\sum_{i \in \tilde{C}} n_i \theta_0} - 1 \right) + R_3$$

$$= U_C - U_{\tilde{C}} + o_p(n_{\min}^{1/2}).$$

(A.4)

Then, we have

$$\sqrt{n_{\min}}(e^{\hat{\alpha}_C} - 1) \overset{D}{\rightarrow} N(0, A_{CC} + A_{\tilde{C}C}).$$

For any $C_j, C_k \in \mathcal{C}$ but $C_j \neq C_k$, we have

$$\text{Cov}[\sqrt{n_{\min}}(e^{\hat{\alpha}_{C_j}} - 1), \sqrt{n_{\min}}(e^{\hat{\alpha}_{C_k}} - 1)] = \text{Cov}(U_{C_j} - U_{\tilde{C_j}}, U_{C_k} - U_{\tilde{C_k}})$$

$$= A_{C_j C_k} - A_{C_j \tilde{C}_k} - A_{\tilde{C}_j C_k} + A_{\tilde{C}_j \tilde{C}_k}.$$ 

Therefore,

$$\sqrt{n_{\min}} \begin{pmatrix} e^{\hat{\alpha}_{C_j} - 1} \\ e^{\hat{\alpha}_{C_k} - 1} \\ \vdots \\ e^{\hat{\alpha}_{C_l} - 1} \end{pmatrix} \overset{D}{\rightarrow} N(0, \Sigma),$$

where $\Sigma$ is given by Eq. (A.3). The conclusion is implied by using the $\Delta$-method if we take the multivariate transformation $g(t_1, t_2, \ldots, t_l) = (\log t_1, \log t_2, \ldots, \log t_l)$. □
Proof of Theorem 1. By using method of empirical process van der Vaart (1998, p. 260), the conclusion is directly drawn from Lemmas 3 and 4 because $C$ is a finite set.

A.2. Proofs for binomial distribution with logistic link

The ideas of the proofs for binomial distribution with the logistic link as well as the probit or complementary link is exactly the same as those of the proofs for Poisson distribution. In this section, we display the proofs for the logistic link and omit the proofs for the probit and complementary links.

If $Y_i$, $i = 1, \ldots, m$, are independent binomial random variables and $g(t) = \log(t/(1 - t))$ is the logistic link, then the log-likelihood function under Model (11) is

$$\ell_C(\alpha, \beta) = \sum_{i \in C} n_i Y_i! (n_i - Y_i)! + \sum_{i = 1}^m Y_i \left( \sum_{j = 1}^p x_j \beta_j + \alpha_C \right) - \sum_{i = 1}^m n_i \log \left( 1 + e^{\sum_{j = 1}^p x_j \beta_j + \alpha_C} \right).$$

Then, the likelihood equations are

$$\frac{\partial \ell_C}{\partial \alpha_C} = 0 \Rightarrow \sum_{i \in C} Y_i = e^{\alpha_C} \sum_{i \in C} \frac{n_i e^{\sum_{j = 1}^p x_j \beta_j}}{1 + e^{\sum_{j = 1}^p x_j \beta_j + \alpha_C}}$$

and

$$\frac{\partial \ell_C}{\partial \beta_k} = 0 \Rightarrow \sum_{i = 1}^m Y_i x_{ik} = e^{\alpha_C} \sum_{i \in C} \frac{n_i e^{\sum_{j = 1}^p x_j \beta_j}}{1 + e^{\sum_{j = 1}^p x_j \beta_j + \alpha_C}} + \sum_{i \notin C} \frac{n_i e^{\sum_{j = 1}^p x_j \beta_j}}{1 + e^{\sum_{j = 1}^p x_j \beta_j + \alpha_C}}$$

for $k = 1, \ldots, p$, where $\hat{Y}_C = \sum_{i \in C} \hat{Y}_C$. Taking $k = 1$ in the Eq. (A.6), we have

$$\sum_{i \notin C} Y_i = \sum_{i \notin C} \frac{n_i e^{\sum_{j = 1}^p x_j \beta_j}}{1 + e^{\sum_{j = 1}^p x_j \beta_j + \alpha_C}}.$$  \hspace{1cm} (A.6)

Denote

$$R_1 = \left[ \frac{\sum_{i \in C} n_i e^{\sum_{j = 1}^p x_j \beta_j}}{1 + e^{\sum_{j = 1}^p x_j \beta_j + \alpha_C}} \right] / \sum_{i \in C} n_i \theta_{i0},$$

$$R_2 = \left[ \frac{\sum_{i \notin C} n_i e^{\sum_{j = 1}^p x_j \beta_j}}{1 + e^{\sum_{j = 1}^p x_j \beta_j + \alpha_C}} \right] / \sum_{i \notin C} n_i \theta_{i0}$$

and

$$R_3 = R_1 - R_2.$$

Then subtract (A.6) from (A.5), we have

$$e^{\alpha_C} - 1 = R_1 \left( \frac{Y_C}{\sum_{i \in C} n_i \theta_{i0}} - 1 \right) + R_2 \left( \frac{Y_C}{\sum_{i \notin C} n_i \theta_{i0}} - 1 \right) + R_3$$

which can be used in the same way as the proofs of the Poisson case in Eq. (A.4). Eq. (A.7) is used to complete the proof for the binomial case in Theorem 1. \hspace{1cm} \Box

References