Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH Open Access

Consistency of bootstrap approximation to the null distributions of local spatial statistics with application to house price analysis

Chang-Lin Mei[1*], Shou-Fang Xu[2,3] and Feng Chen[2]

*Correspondence:
clmei@xpu.edu.cn
[1]School of Science, Xi'an Polytechnic University, Xi'an, China
Full list of author information is available at the end of the article

Abstract

With the increasing availability of spatially extensive geo-referenced data, much attention has been paid to the use of local statistics to identify local patterns of spatial association, in which the null distributions of local statistics play an essential role in the related statistical inference. As a powerful tool to approximate the distribution of a statistic, the bootstrap method is used in this paper to derive null distributions of the commonly used local spatial statistics including local Getis and Ord's G_i, Moran's I_i and Geary's c_i. Strong consistency of the bootstrap approximation to the null distributions of the statistics is proved under some mild conditions, and the Boston housing price data are analyzed to demonstrate the application of the theoretical results.

MSC: Primary 62G09; secondary 62G20

Keywords: Local spatial statistic; Bootstrap; Strong consistency; Kolmogorov distance; Mallows distance

1 Introduction

Exploration of spatial association has long been recognized as an important issue in spatial data analysis. With the increasing availability of spatially extensive geo-referenced data and due to the geological and geographical diversity on a large region, a global structure of spatial association is no longer a realistic assumption for such a data set. Therefore, much attention has been paid to the use of local statistics to identify local patterns of spatial association. The most popular local spatial statistics are perhaps Getis and Ord's G_i [11, 18] and Anselin's LISAs [1]. Since their inception, these local statistics have been applied to a variety of fields for spatial data analysis (see, for example, [9, 10, 14, 24]).

In order to test for significance of local spatial association at a reference location, it is essential to derive the null distribution of the local statistics. Normal distributions have been used to approximate the null distributions of some local spatial statistics such as local Getis and Ord's G_i, Moran's I_i and Geary's c_i (see, for example, [1, 11, 18]). However, many empirical studies have shown that this approximation is sometimes problematic [1, 4, 5, 27]. Based on the distributional theory of quadratic forms in normal vari-

ables, some improved methods have been developed under the assumption that the spatial data are drawn from a normally distributed population (see, for example, [5, 13, 20–22]). Nevertheless, this assumption might be invalid for some real-world data sets. With the computation power of modern computers, the randomized permutation method, a resampling procedure that randomly relocates the data over the locations, is frequently employed to approximate the null distributions of local spatial statistics (see, for example, [1, 12, 17]). Recently, Yan et al. [26] suggested a bootstrap method, originally proposed by Efron [8], to approximate the null distributions of the spatio-temporal versions of local Getis and Ord's G_i, Moran's I_i and Geary's c_i. They showed by simulations that both the bootstrap and the randomized permutation methods can accurately approximate the null distributions of the local statistics while the bootstrap method seems more efficient than the randomized permutation method in terms of computational time. However, the theoretical validity of the bootstrap approximation remains to be investigated.

The main objective of this paper is to theoretically investigate the validity of the bootstrap approximation to the null distributions of local Getis and Ord's G_i, Moran's I_i and Geary's c_i. Under some mild conditions, we proved that the bootstrap approximation is strongly consistent in terms of the Kolmogorov distance on the space of distribution functions. Moreover, the Monte Carlo implementation of the bootstrap approximation for statistical inference is given in detail by a case study of the Boston housing price in order to demonstrate application of the theoretical results.

The remainder of this paper is organized as follows: the main results are presented in the next section, and their proofs are given in Sect. 3. As an application example of the theoretical results, the Boston housing price data are analyzed in Sect. 4. The paper is then ended with a brief summary.

2 Main results

Let $F(x)$ be the population distribution and s be the coordinate of a geographical location. Given n locations s_i $(i = 1, 2, \ldots, n)$, let $W = (w_{ij}(d))_{n\times n}$ be the symmetric spatial linkage matrix determined by the underlying spatial structure of the n locations or geographical units, where d is a pre-specified distance threshold and $w_{ij}(d)$ $(j = 1, 2, \ldots, n)$ are positive for all s_j's within distance d of the location s_i excluding $s_j = s_i$, and are zero for other s_j's. Generally, the binary values, zero and one, are assigned to $w_{ij}(d)$ $(j = 1, 2, \ldots, n)$ according to the above rule. At each location s_j, draw independently X_j from the population distribution $F(x)$, forming an independent and identically distributed (i.i.d.) sample $(X_1, X_2, \ldots, X_n)$ with X_j located at s_j $(j = 1, 2, \ldots, n)$.

Given a reference location s_i, after re-scaling and/or re-centering, the local Getis and Ord's G_i [11], the local Moran's I_i and Geary's c_i [1] are, respectively, of the forms

$$G_i(d) = \frac{n-1}{W_{in}} \frac{\sum_{j=1}^{n} w_{ij}(d)(X_j - \bar{X})}{\sum_{j\neq i} X_j}, \tag{1}$$

$$I_i(d) = \frac{n}{W_{in}} \frac{(X_i - \bar{X}) \sum_{j=1}^{n} w_{ij}(d)(X_j - \bar{X})}{\sum_{j=1}^{n} (X_j - \bar{X})^2}, \tag{2}$$

and

$$c_i(d) = \frac{n}{W_{in}^2} \frac{\sum_{j=1}^n w_{ij}(d)(X_i - X_j)^2}{\sum_{j=1}^n (X_j - \bar{X})^2}, \tag{3}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $W_{in} = \sqrt{\sum_{j=1}^n w_{ij}(d)}$.

Remark 1 For $G_i(d)$, it is a natural assumption that $E(X_j) = \mu \neq 0$. Moreover, we modify the numerator in the $G_i(d)$ statistic as $X_j - \bar{X}$ instead of X_j in its original form to facilitate the forthcoming proof of the asymptotic property. This modification does not change the interpretation of the statistic.

Let F_n denote the empirical distribution of the sample $(X_1, X_2, \ldots, X_n)$, that is,

$$F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{\{X_i \leq x\}},$$

where $\mathbb{I}_{\{A\}}$ is the indicator function of the event A. Let $(X_1^*, X_2^*, \ldots, X_n^*)$ be the bootstrap sample drawn from $F_n(x)$ with replacement and be located at $(s_1, s_2, \ldots, s_n)$. The bootstrap scenarios of the local Getis and Ord's G_i, Moran's I_i and Geary's c_i are, respectively,

$$G_i^*(d) = \frac{n-1}{W_{in}} \frac{\sum_{j=1}^n w_{ij}(d)(X_j^* - \bar{X}^*)}{\sum_{j\neq i} X_j^*}, \tag{4}$$

$$I_i^*(d) = \frac{n}{W_{in}} \frac{\sum_{j=1}^n w_{ij}(d)(X_i^* - \bar{X}^*)(X_j^* - \bar{X}^*)}{\sum_{j=1}^n (X_j^* - \bar{X}^*)^2}, \tag{5}$$

and

$$c_i^*(d) = \frac{n}{W_{in}^2} \frac{\sum_{j=1}^n w_{ij}(d)(X_i^* - X_j^*)^2}{\sum_{j=1}^n (X_j^* - \bar{X}^*)^2}, \tag{6}$$

where $\bar{X}^* = \frac{1}{n}\sum_{i=1}^n X_i^*$.

Throughout this paper, we use the notations P, E and Var to indicate the probability, expectation and variance calculated under $F(x)$ and the notations P^*, E^* and Var* to represent those computed under $F_n(x)$. In what follows, we first introduce the consistency definition of bootstrap approximation to the distribution of a statistic and then give the main results of this article.

Definition 1 ([7], Chap. 29) Let F and G be two distributions on a sample space $\mathscr{X}$ and $\rho(F, G)$ be a metric on the space of distribution functions. Let $(X_1, X_2, \ldots, X_n)$ be i.i.d. random variables with the common distribution F. For a given statistic $T = T(X_1, \ldots, X_n; F)$, let $H_n(x) = P(T(X_1, X_2, \ldots, X_n; F) \leq x)$ and $H_n^*(x) = P^*(T(X_1^*, X_2^*, \ldots, X_n^*; F_n) \leq x)$ be the distribution function of T and the bootstrap distribution function of $T^* = T(X_1^*, X_2^*, \ldots, X_n^*; F_n)$, respectively. We say that the bootstrap approximation for T is weakly consistent under ρ if $\rho(H_n, H_n^*) \xrightarrow{P} 0$ as $n \to \infty$, where $\xrightarrow{P}$ denotes convergence in probability; we say that the bootstrap approximation for T is strongly consistent under ρ if $\rho(H_n, H_n^*) \xrightarrow{a.s.} 0$ as $n \to \infty$, where $\xrightarrow{a.s.}$ denotes convergence for almost all sample sequences of $X_1, X_2, \ldots$.

Several metrics such as the Kolmogorov distance and the Mallows distance can be employed to measure the consistency of the bootstrap approximation. The Kolmogorov distance defined by

$$K(F,G) = \sup_{x\in R}\left|F(x) - G(x)\right|$$

is commonly used, where $R = (-\infty, +\infty)$. In this paper, the Kolmogorov distance is mainly used to investigate the strong consistency of the bootstrap approximation for the local Getis and Ord's G_i, Moran's I_i and Geary's c_i and the main results are summarized in the following theorems.

Theorem 1 *Let $W = (w_{ij}(d))_{n\times n}$ be the binary spatial linkage matrix of the geographical locations s_j $(j = 1, 2, \ldots, n)$ and $W_{in} = \sqrt{\sum_{j=1}^{n} w_{ij}(d)}$. Let $(X_1, X_2, \ldots, X_n)$ be an i.i.d. sample drawn from a continuous distribution F with non-zero mean μ and positive variance σ^2. Given a reference location s_i, if $\frac{1}{n}W_{in}^2 \to 0$ as $n \to \infty$, then the bootstrap approximation for $G_i(d)$ is strongly consistent under the Kolmogorov distance. That is,*

$$\sup_{x\in R}\left|P^*\left(G_i^*(d) \le x\right) - P\left(G_i(d) \le x\right)\right| \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.$$

Theorem 2 *Let $W = (w_{ij}(d))_{n\times n}$ be the binary spatial linkage matrix of the geographical locations s_j $(j = 1, 2, \ldots, n)$ and $W_{in} = \sqrt{\sum_{j=1}^{n} w_{ij}(d)}$. Let $(X_1, X_2, \ldots, X_n)$ be an i.i.d. sample drawn from a continuous distribution F with mean μ and positive variance σ^2. Given a reference location s_i, if $\frac{1}{n}W_{in}^2 \to 0$ as $n \to \infty$, then the bootstrap approximation for $I_i(d)$ is strongly consistent under the Kolmogorov distance. That is,*

$$\sup_{x\in R}\left|P^*\left(I_i^*(d) \le x\right) - P\left(I_i(d) \le x\right)\right| \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.$$

Theorem 3 *Let $W = (w_{ij}(d))_{n\times n}$ be the binary spatial linkage matrix of the geographical locations s_j $(j = 1, 2, \ldots, n)$ and $W_{in} = \sqrt{\sum_{j=1}^{n} w_{ij}(d)}$. Let $(X_1, X_2, \ldots, X_n)$ be an i.i.d. sample drawn from a continuous distribution F with mean μ and positive variance σ^2. Given a reference location s_i, the bootstrap approximation for $c_i(d)$ is strongly consistent under the Kolmogorov distance. That is,*

$$\sup_{x\in R}\left|P^*\left(c_i^*(d) \le x\right) - P\left(c_i(d) \le x\right)\right| \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.$$

3 Proofs of the main results

3.1 Preliminaries and lemmas

To prove the theorems, the Mallows distance (see, for example, [3, 15, 16]) will be used because of its interesting properties relating to the Kolmogorov distance. Let $\mathscr{F}_p$ be the set of distribution functions F with $\int_{-\infty}^{\infty} |x|^p \, dF(x) < \infty$. For $F, G \in \mathscr{F}_p$, the Mallows distance between F and G is defined as

$$d_p(F,G) = \inf_{(X,Y)} \left\{\left(E|X - Y|^p\right)^{\frac{1}{p}}\right\},$$

where $1 \le p < \infty$ and the infimum is taken over the pairs (X, Y) with the marginal distribution functions of X and Y being F and G, respectively. Throughout this paper, we also write $d_p(F, G)$ and $[d_p(F, G)]^2$ as $d_p(X, Y)$ and $d_p^2(X, Y)$, respectively, for the ease of interpretation.

Lemma 1 ([23], p. 12) *Let $X_1, X_2, \dots$ be a random variable sequence and X be a random variable with a continuous distribution function. If X_n converges to X in distribution, which we denote $X_n \rightsquigarrow X$, then*

$$\sup_{x \in R} \left| P(X_n \le x) - P(X \le x) \right| \to 0 \quad \text{as } n \to \infty.$$

Lemma 2 ([3]) *Let $G_n \in \mathscr{F}_p$ and $G \in \mathscr{F}_p$. Then $d_p(G_n, G) \to 0$ as $n \to \infty$ if and only if both of the following conditions hold*:

(1) $G_n \to G$ *weakly as* $n \to \infty$,

(2) $\lim_{n \to \infty} \int_{-\infty}^{\infty} |x|^p \, dG_n(x) = \int_{-\infty}^{\infty} |x|^p \, dG(x)$.

Remark 2 Let the distribution functions of X_n and X be G_n and G, respectively. Lemma 2 means that G_n converges to G in the Mallows distance d_p if and only if $X_n \rightsquigarrow X$ and $E|X_n|^p \to E|X|^p$.

Lemma 3 *Let $X_1, X_2, \dots$ be an i.i.d. random variable sequence with the common distribution function $F \in \mathscr{F}_p$. Let F_n be the empirical distribution function of $(X_1, X_2, \dots, X_n)$. Then*

$$d_p(F_n, F) \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty,$$

where $\xrightarrow{a.s.}$ means that $d_p(F_n, F) \to 0$ for almost all sample sequences of $X_1, X_2, \dots$.

Proof By Lemma 2, it is sufficient to prove $F_n \to F$ weakly and $E|X_n|^p \to E|X|^p$. Let $Y_i = \mathbb{I}_{\{X_i \le x\}}$ $(i = 1, 2, \dots, n)$. Since $(Y_1, Y_2, \dots, Y_n)$ are i.i.d. random variables, we know from the strong law of large numbers that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} Y_i = E(Y_i) = P(X_i \le x) = F(x), \quad \text{a.s.},$$

which indicates $F_n \to F$ weakly. Similarly, $E|X_n|^p \to E|X|^p$ can be obtained by using the strong law of large numbers for $(|X_1|^p, |X_2|^p, \dots, |X_n|^p)$. □

Lemma 4 ([3]) *Let $(X_1, X_2, \dots, X_n)$ and $(Y_1, Y_2, \dots, Y_n)$ be two sets of independent random variables with their distribution functions belonging to $\mathscr{F}_p$. Then, for constants a_i $(1 \le i \le n)$, we have*

$$d_p\left(\sum_{i=1}^{n} a_i X_i, \sum_{i=1}^{n} a_i Y_i \right) \le \sum_{i=1}^{n} |a_i| d_p(X_i, Y_i).$$

Remark 3 The key for proving this lemma is the use of the Minkowski's inequality (see Lemma 8.6 in Bickel and Freedman [3] for the details), which does not need the independence condition among the two sets of the random variables. Therefore, the independence assumption on $(X_1, X_2, \ldots, X_n)$ as well as on $(Y_1, Y_2, \ldots, Y_n)$ is indeed not indispensable for guaranteeing the conclusion of the lemma.

Lemma 5 *Let $X, X_1, X_2, \ldots$ be a sequence of random variables with their distribution functions belonging to $\mathscr{F}_p$. If $d_p(X_n, X) \to 0$ as $n \to \infty$, then*

$$d_{p/2}\big(X_n^2, X^2\big) \to 0 \quad \text{as } n \to \infty.$$

Proof The conditions imply that the distribution functions of $X^2, X_1^2, X_2^2, \ldots$ belong to $\mathscr{F}_q$. From Lemma 2, we have (i) $X_n \rightsquigarrow X$; and (ii) $E(|X_n|^p) \to E(|X|^p)$. The continuous mapping theorem ([23], p.7) together with (i) yields $(iii) X_n^2 \rightsquigarrow X^2$. The lemma is then proved according to (ii), (iii) and Lemma 2. □

Lemma 6 *Let $X_1, X_2, \ldots$ be an i.i.d. random variable sequence drawn from F with finite variance σ^2. Let F_n and $(X_1^*, X_2^*, \ldots, X_n^*)$ be the empirical distribution function and the bootstrap sample of $(X_1, X_2, \ldots, X_n)$, respectively. Then, for almost all sample sequences of $X_1, X_2, \ldots$,*

$$\frac{1}{n}\sum_{i=1}^{n}\big(X_i^* - \bar{X}^*\big)^2 \xrightarrow{a.s.} \sigma^2 \quad \text{as } n \to \infty.$$

Proof The condition $\sigma^2 < \infty$ implies that $E(X_i) \triangleq \mu$ exists. By the strong law of large numbers, we have, for almost all sample sequences of $X_1, X_2, \ldots$,

$$\bar{X}^* \xrightarrow{a.s.} \mu \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}\big(X_i^*\big)^2 \xrightarrow{a.s.} \mu^2 + \sigma^2 \quad \text{as } n \to \infty.$$

Then the desired result can be proved by the continuous mapping theorem. □

Lemma 7 *If X_n and Y_n are independent random variables for each n, then $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$ imply that $(X_n, Y_n) \rightsquigarrow (X, Y)$ with X and Y being independent.*

Proof Because X_n and Y_n are independent random variables for every n, we have $F_{(X_n,Y_n)} = F_{X_n}F_{Y_n}$. It follows from $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$ that $F_{(X_n,Y_n)} \to F_X F_Y$ for all continuous points of $F_X F_Y$. Then the lemma is proved. □

3.2 Proofs of the theorems

In the proofs of the theorems, the following two cases will be separately considered because the proof ways are essentially different for the two cases.

Case 1 Suppose that $W_{in} = \sqrt{\sum_{j=1}^{n} w_{ij}(d)} < \infty$ as $n \to \infty$, which means that the number of observations within d-distance neighborhood of the reference location s_i will be fixed when n is large enough. For a local spatial statistic, this case is possible if the newly coming observations are all placed outside the d-distance neighborhood of the reference location s_i after n reaches some finite integer, say, n_0.

Case 2 Assume $W_{in} \to \infty$ as $n \to \infty$, which implies that the number of observations within distance d of the reference location s_i goes to infinity as $n \to \infty$.

Proof of Theorem 1 Note that

$$E^*\left(\frac{1}{n-1}\sum_{j\neq i} X_j^*\right) = E^*\left(X_1^*\right) = \bar{X}.$$

Since $\bar{X} \xrightarrow{a.s.} \mu$ as $n \to \infty$, we have

$$\frac{1}{n-1}\sum_{j\neq i} X_j^* \xrightarrow{a.s.} \mu \quad \text{for almost all sample sequences of } X_1, X_2, \ldots. \tag{7}$$

Furthermore, the numerators of $G_i(d)$ and $G_i^*(d)$ can be, respectively, expressed as

$$\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)(X_j - \bar{X}) = \frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)(X_j - \mu) + W_{in}(\mu - \bar{X}) \tag{8}$$

and

$$\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)\left(X_j^* - \bar{X}^*\right) = \frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)\left(X_j^* - \bar{X}\right) + W_{in}\left(\bar{X} - \bar{X}^*\right). \tag{9}$$

For any $\varepsilon > 0$, by the Chebyshev inequality and the assumption that $\frac{1}{n}W_{in}^2 \to 0$ as $n \to \infty$, we obtain

$$P\left(\left|W_{in}(\mu - \bar{X})\right| \geq \varepsilon\right) \leq \frac{\operatorname{Var}(W_{in}(\mu - \bar{X}))}{\varepsilon^2} = \frac{W_{in}^2\sigma^2}{n\varepsilon^2} \to 0 \quad \text{as } n \to \infty,$$

which implies

$$W_{in}(\mu - \bar{X}) \xrightarrow{P} 0 \quad \text{as } n \to \infty. \tag{10}$$

Similarly, we have

$$P^*\left(\left|W_{in}\left(\bar{X} - \bar{X}^*\right)\right| \geq \varepsilon\right) \leq \frac{\operatorname{Var}^*(W_{in}(\bar{X} - \bar{X}^*))}{\varepsilon^2} = \frac{W_{in}^2}{n\varepsilon^2}\operatorname{Var}^*\left(X_1^*\right) = \frac{W_{in}^2}{n\varepsilon^2}S_n^2,$$

where $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Since $S_n^2 \xrightarrow{a.s.} \sigma^2 < \infty$ according to the strong law of large numbers, we have, for almost all sample sequences of $X_1, X_2, \ldots,$

$$P^*\left(\left|W_{in}\left(\bar{X} - \bar{X}^*\right)\right| \geq \varepsilon\right) \to 0 \quad \text{as } n \to \infty,$$

which implies

$$W_{in}\left(\bar{X} - \bar{X}^*\right) \xrightarrow{P^*} 0 \quad \text{for almost all sample sequences of } X_1, X_2, \ldots. \tag{11}$$

In Case 1, from Eqs. (8) and (10) and the Slutsky theorem, we have

$$\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)(X_j-\bar{X}) \rightsquigarrow \frac{1}{W_{in_0}}\sum_{j=1}^{n_0} w_{ij}(d)(X_j-\mu) \triangleq Z_0 \quad \text{as } n\to\infty, \tag{12}$$

where $W_{in_0} = \sqrt{\sum_{j=1}^{n_0} w_{ij}(d)}$. Similarly, from Eqs. (9) and (11), it can be inferred that $\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)(X_j^*-\bar{X}^*)$ and $\frac{1}{W_{in_0}}\sum_{j=1}^{n_0} w_{ij}(d)(X_j^*-\bar{X})$ have the same limiting distribution for almost all sample sequences of $X_1, X_2, \ldots$. Moreover, according to Lemmas 2, 3 and 4, we obtain

$$\begin{aligned} &d_1\left(\frac{1}{W_{in_0}}\sum_{j=1}^{n_0} w_{ij}(d)\big(X_j^*-\bar{X}\big), \frac{1}{W_{in_0}}\sum_{j=1}^{n_0} w_{ij}(d)(X_j-\mu)\right) \\ &\quad\le W_{in_0} d_1\big(X_j^*-\bar{X}, X_j-\mu\big) \\ &\quad\le W_{in_0}\big[d_1\big(X_j^*, X_j\big) + d_1(\bar{X},\mu)\big] \xrightarrow{a.s.} 0, \end{aligned}$$

which implies that the distribution of $\frac{1}{W_{in_0}}\sum_{j=1}^{n_0} w_{ij}(d)(X_j^*-\bar{X})$ converges to the distribution of Z_0. Therefore, for almost all sample sequences of $X_1, X_2, \ldots$, we have

$$\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)\big(X_j^*-\bar{X}^*\big) \rightsquigarrow Z_0 \quad \text{as } n\to\infty. \tag{13}$$

On the other hand, let $G(d) \triangleq \frac{Z_0}{\mu}$ where $\mu = E(X_i) \neq 0$. From the Slutsky theorem, Eq. (12) and the fact that $\frac{1}{n-1}\sum_{j\neq i} X_j \xrightarrow{a.s.} \mu$ as $n\to\infty$, we have $G_i(d) \rightsquigarrow G(d)$. Then, according to Lemma 1 and noting that the distribution function of $G(d)$ is continuous, we have

$$\sup_{x\in R}\big|P\big(G_i(d)\le x\big) - P\big(G(d)\le x\big)\big| \to 0 \quad \text{as } n\to\infty.$$

Similarly, from Eqs. (7) and (13), we have

$$\sup_{x\in R}\big|P^*\big(G_i^*(d)\le x\big) - P\big(G(d)\le x\big)\big| \xrightarrow{a.s.} 0 \quad \text{as } n\to\infty.$$

Therefore, the above two equations and the triangle inequality yields the conclusion of the theorem.

In Case 2, given an n, suppose that there are k_n observation locations within distance d of s_i, leading to $W_{in} = \sqrt{k_n}$ and $k_n\to\infty$ as $n\to\infty$. Without loss of generality, let $(X_1, X_2, \ldots, X_{k_n})$ locate within distance d of s_i. Note that $X_j-\mu$ $(j=1,\ldots,k_n)$ are i.i.d. random variables with finite variance σ^2 and $k_n\to\infty$ as $n\to\infty$. Therefore, according to the central limit theorem, we have

$$\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)(X_j-\mu) = \frac{1}{\sqrt{k_n}}\sum_{j=1}^{k_n}(X_j-\mu) \rightsquigarrow Z \quad \text{as } n\to\infty,$$

where Z stands for a random variable distributed as the normal distribution $N(0,\sigma^2)$. It therefore follows from Eqs. (8) and (10) and the Slutsky theorem that

$$\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)(X_j-\bar{X}) \rightsquigarrow Z \quad \text{as } n\to\infty. \tag{14}$$

Similarly, because $X_j^*-\bar{X}$ $(j=1,\dots,k_n)$ are conditionally i.i.d. random variables with variance $S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ and $k_n\to\infty$ as $n\to\infty$, then, according to the central limit theorem and noting $S_n^2\xrightarrow{a.s.}\sigma^2$ as $n\to\infty$, we have, for almost all sample sequences of $X_1,X_2,\dots$,

$$\frac{1}{\sqrt{k_n}}\sum_{j=1}^{k_n}\left(X_j^*-\bar{X}\right) \rightsquigarrow Z \quad \text{as } n\to\infty.$$

This, together with Eqs. (9) and (11) and the Slutsky theorem, yields

$$\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)\left(X_j^*-\bar{X}^*\right) \rightsquigarrow Z \quad \text{as } n\to\infty \tag{15}$$

for almost all sample sequences of $X_1,X_2,\dots$. Let $G(d)\triangleq\frac{Z}{\mu}$. With a similar derivation to that in Case 1, the theorem is then proved in this case. □

Proof of Theorem 2 Notice that the numerator of $I_i(d)$ can be expressed as

$$\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)(X_i-\bar{X})(X_j-\bar{X}) \triangleq T_1+T_2+T_3, \tag{16}$$

where

$$T_1=W_{in}(\bar{X}-\mu)(\bar{X}-X_i); \qquad T_2=\frac{1}{W_{in}}(\mu-\bar{X})\sum_{j=1}^{n} w_{ij}(d)(X_j-\mu);$$

$$T_3=\frac{1}{W_{in}}(X_i-\mu)\sum_{j=1}^{n} w_{ij}(d)(X_j-\mu).$$

Firstly, from Eq. (10) and $\bar{X}-X_i\xrightarrow{a.s.}\mu-X_i$, we have $T_1\xrightarrow{P}0$ as $n\to\infty$.

Secondly, as mentioned in the proof of Theorem 1, $\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)(X_j-\mu)$ converges to Z_0 and Z in distribution as $n\to\infty$ in Cases 1 and 2, respectively. Therefore, from $\mu-\bar{X}\xrightarrow{a.s.}0$ and the Slutsky theorem, we have $T_2\xrightarrow{P}0$ as $n\to\infty$ in both cases.

Finally, because $(x,y)\mapsto yx$ is a continuous mapping, then, by Lemma 7 and the result that $X_i-\mu$ is independent from $\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)(X_j-\mu)$, we have $T_3\rightsquigarrow(X_i-\mu)Z_0$ and $T_3\rightsquigarrow(X_i-\mu)Z$ as $n\to\infty$ in Cases 1 and 2, respectively.

By the Slutsky theorem and Eq. (16), we obtain, as $n\to\infty$,

$$\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)(X_i-\bar{X})(X_j-\bar{X}) \rightsquigarrow (X_i-\mu)Z_0$$

in Case 1 and

$$\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)(X_i-\bar{X})(X_j-\bar{X}) \rightsquigarrow (X_i-\mu)Z$$

in Case 2.

Let $I(d) \triangleq \frac{(X_i-\mu)Z_0}{\sigma^2}$ and $I(d) \triangleq \frac{(X_i-\mu)Z}{\sigma^2}$ in Cases 1 and 2, respectively. Since $\frac{1}{n}\sum_{j=1}^{n}(X_j-\bar{X})^2 \xrightarrow{a.s.} \sigma^2$ as $n\to\infty$, we know that $I_i(d) \rightsquigarrow I(d)$ according to the Slutsky theorem. Therefore, Lemma 1 and the continuity of the distribution function of $I(d)$ guarantee that

$$\sup_{x\in R}\left|P\big(I_i(d)\le x\big)-P\big(I(d)\le x\big)\right| \to 0 \quad \text{as } n\to\infty.$$

According to the triangle inequality, to prove Theorem 2, it is sufficient to prove

$$\sup_{x\in R}\left|P^*\big(I_i^*(d)\le x\big)-P\big(I(d)\le x\big)\right| \xrightarrow{a.s.} 0 \quad \text{as } n\to\infty.$$

In a similar way to that in dealing with the quantity of the left-hand side in Eq. (16), we rewrite the numerator of $I_i^*(d)$ as

$$\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)\big(X_i^*-\bar{X}^*\big)\big(X_j^*-\bar{X}^*\big) \triangleq T_1^*+T_2^*+T_3^*, \tag{17}$$

where

$$T_1^* = W_{in}\big(\bar{X}^*-\bar{X}\big)\big(\bar{X}^*-X_i^*\big); \qquad T_2^* = \frac{1}{W_{in}}\big(\bar{X}-\bar{X}^*\big)\sum_{j=1}^{n} w_{ij}(d)\big(X_j^*-\bar{X}\big);$$

$$T_3^* = \frac{1}{W_{in}}\big(X_i^*-\bar{X}\big)\sum_{j=1}^{n} w_{ij}(d)\big(X_j^*-\bar{X}\big).$$

First of all, we obtain $T_1^* \xrightarrow{P^*} 0$ as $n\to\infty$ according to Eq. (11) and $\bar{X}^*-X_i^* \xrightarrow{a.s.} \mu-X_i^*$ as $n\to\infty$ for almost all sample sequences of $X_1, X_2, \ldots$.

Then, for almost all sample sequences of $X_1, X_2, \ldots$, we have

$$\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)\big(X_j^*-\bar{X}\big) \rightsquigarrow Z_0 \quad \text{as } n\to\infty$$

in Case 1 and

$$\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)\big(X_j^*-\bar{X}\big) \rightsquigarrow Z \quad \text{as } n\to\infty$$

in Case 2. Furthermore, it follows from the Slutsky theorem and $\bar{X}-\bar{X}^* \xrightarrow{a.s.} 0$ as $n\to\infty$ that $T_2^* \xrightarrow{P^*} 0$ as $n\to\infty$ in both cases.

Finally, it is known that

$$d_1\big(X_i^* - \bar{X}, X_i - \mu\big) \le d_1\big(X_i^*, X_i\big) + d_1(\bar{X}, \mu) \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty,$$

which implies $X_i^* - \bar{X} \rightsquigarrow X_i - \mu$ as $n \to \infty$. Then, according to Lemma 7 and the result that $X_i^* - \bar{X}$ is conditionally independent to $\frac{1}{W_{in}} \sum_{j=1}^n w_{ij}(d)(X_j^* - \bar{X})$, we obtain $T_3^* \rightsquigarrow (X_i - \mu)Z_0$ and $T_3^* \rightsquigarrow (X_i - \mu)Z$ as $n \to \infty$ in Cases 1 and 2, respectively.

According to the Slutsky theorem, it follows from Lemma 6 and Eq. (17) that $I_i^*(d) \rightsquigarrow I(d)$ as $n \to \infty$ in both cases. Noting the continuity of the distribution function of $I(d)$ and using Lemma 1 and the triangle inequality, Theorem 2 is then proved. □

Proof of Theorem 3 In Case 1, since $W_{in} = \sqrt{\sum_{j=1}^n w_{ij}(d)} < \infty$ as $n \to \infty$, we can write $W_{in} = W_{in_0} = \sqrt{\sum_{j=1}^{n_0} w_{ij}(d)}$ for some positive integer n_0. According to the triangle inequality, the Hölder inequality and Lemma 3, we have

$$\begin{aligned} &d_1\big(X_i^* X_j^*, X_i X_j\big) \\ &\quad \le d_1\big(X_i^* X_j^*, X_i^* X_j\big) + d_1\big(X_i^* X_j, X_i X_j\big) \\ &\quad \le E^*\big|X_i^*\big| d_1\big(X_j^*, X_j\big) + E|X_j| d_1\big(X_i^*, X_i\big) \xrightarrow{a.s.} 0. \end{aligned}$$

Then it follows from Lemmas 4 and 5 that

$$\begin{aligned} &d_1\left(\frac{1}{W_{in_0}^2} \sum_{j=1}^{n_0} w_{ij}(d)\big(X_i^* - X_j^*\big)^2, \frac{1}{W_{in_0}^2} \sum_{j=1}^{n_0} w_{ij}(d)(X_i - X_j)^2\right) \\ &\quad \le d_1\big(\big(X_i^*\big)^2 - 2X_i^* X_j^* + \big(X_j^*\big)^2, X_i^2 - 2X_i X_j + X_j^2\big) \\ &\quad \le 2\big(d_1\big(\big(X_i^*\big)^2, X_i^2\big) + d_1\big(X_i^* X_j^*, X_i X_j\big)\big) \xrightarrow{a.s.} 0, \end{aligned}$$

which implies that both $\frac{1}{W_{in}^2} \sum_{j=1}^n w_{ij}(d)(X_i^* - X_j^*)^2$ and $\frac{1}{W_{in}^2} \sum_{j=1}^n w_{ij}(d)(X_i - X_j)^2$ converge to $\frac{1}{W_{in_0}^2} \sum_{j=1}^{n_0} w_{ij}(d)(X_i - X_j)^2 \triangleq T$ in distribution as $n \to \infty$.

Let $c(d) = \frac{T}{\sigma^2}$. From the fact that $S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \xrightarrow{a.s.} \sigma^2$ as $n \to \infty$ and the Slutsky theorem, we have $c_i(d) \rightsquigarrow c(d)$ as $n \to \infty$. According to Lemma 1 and the continuity of the distribution function of $c(d)$, we obtain

$$\sup_{x \in R} \big|P\big(c_i(d) \le x\big) - P\big(c(d) \le x\big)\big| \to 0 \quad \text{as } n \to \infty.$$

Similarly, from Lemma 6, we have

$$\sup_{x \in R} \big|P^*\big(c_i^*(d) \le x\big) - P\big(c(d) \le x\big)\big| \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.$$

Then the theorem is proved by using the triangle inequality.

In Case 2, since $W_{in} \to \infty$ as $n \to \infty$, we can rewrite the numerator of $c_i(d)$ as

$$\frac{1}{W_{in}^2} \sum_{j=1}^n w_{ij}(d)(X_i - X_j)^2 = A + B + C, \tag{18}$$

where

$$A = X_i^2 - 2\mu X_i + \mu^2 + \sigma^2; \qquad B = \frac{1}{W_{in}^2}\sum_{j=1}^{n} w_{ij}(d)\big(X_j^2 - \mu^2 - \sigma^2\big);$$

$$C = -\frac{2X_i}{W_{in}^2}\sum_{j=1}^{n} w_{ij}(d)(X_j - \mu).$$

With the same argument as in the proof of Theorem 1, we have

$$\frac{1}{W_{in}^2}\sum_{j=1}^{n} w_{ij}(d)\big(X_j^2 - \mu^2 - \sigma^2\big) = \frac{1}{k_n}\sum_{j=1}^{k_n}\big(X_j^2 - \mu^2 - \sigma^2\big).$$

According to the strong law of large numbers, we obtain $B \xrightarrow{a.s.} 0$ as $n \to \infty$.

It follows from the Markovian inequality that

$$P\left(\left|\frac{2X_i}{W_{in}}\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}E\left(\frac{2X_i}{W_{in}}\right)^2 = \frac{4(\mu^2 + \sigma^2)}{W_{in}^2\varepsilon^2} \to 0 \quad \text{as } n \to \infty,$$

that is,

$$\frac{2X_i}{W_{in}} \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

Then the Slutsky theorem together with the result that $\frac{1}{W_{in}}\sum_{j=1}^{n} w_{ij}(d)(X_j - \mu) \rightsquigarrow Z$ as $n \to \infty$ guarantees $C \xrightarrow{P} 0$ as $n \to \infty$.

Applying the Slutsky theorem to Eq. (18), we have

$$\frac{1}{W_{in}^2}\sum_{j=1}^{n} w_{ij}(d)(X_i - X_j)^2 \rightsquigarrow A \quad \text{as } n \to \infty.$$

Let $c(d) \triangleq \frac{A}{\sigma^2}$. From the Slutsky theorem and $S_n^2 \xrightarrow{a.s.} \sigma^2$, we obtain $c_i(d) \rightsquigarrow c(d)$. Then, according to Lemma 1 and the assumption that the distribution function of $c(d)$ is continuous, we obtain

$$\sup_{x \in R}\big|P\big(c_i(d) \leq x\big) - P\big(c(d) \leq x\big)\big| \to 0 \quad \text{as } n \to \infty.$$

By the triangle inequality for the Kolmogorov distance, it is then sufficient to prove

$$\sup_{x \in R}\big|P^*\big(c_i^*(d) \leq x\big) - P\big(c(d) \leq x\big)\big| \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty. \tag{19}$$

Similarly, the numerator of $c_i^*(d)$ can be expressed as

$$\frac{1}{W_{in}^2}\sum_{j=1}^{n} w_{ij}(d)\big(X_i^* - X_j^*\big)^2 = A^* + B^* + C^*, \tag{20}$$

where

$$A^* = \left(X_i^*\right)^2 - 2\bar{X}X_i^* + (\bar{X})^2 + S_n^2; \qquad B^* = \frac{1}{W_{in}^2}\sum_{j=1}^{n} w_{ij}(d)\left[\left(X_j^*\right)^2 - (\bar{X})^2 - S_n^2\right];$$

$$C^* = -\frac{2X_i^*}{W_{in}^2}\sum_{j=1}^{n} w_{ij}(d)\left(X_j^* - \bar{X}\right).$$

Firstly, according to Lemmas 3, 4 and 5, we obtain

$$\begin{aligned} &d_1\left(\left(X_i^*\right)^2 - 2\mu X_i^* + \mu^2 + \sigma^2, X_i^2 - 2\mu X_i + \mu^2 + \sigma^2\right) \\ &\quad \le d_1\left(\left(X_i^*\right)^2, X_i^2\right) + 2\mu d_1\left(X_i^*, X_i\right) \xrightarrow{a.s.} 0, \end{aligned}$$

which implies that the distribution of $(X_i^*)^2 - 2\mu X_i^* + \mu^2 + \sigma^2$ converges to the distribution of A. According to the strong law of large numbers, we therefore obtain $A^* \rightsquigarrow A$ as $n \to \infty$ for almost all sample sequences of $X_1, X_2, \ldots$.

Moreover, with the same argument in the proof of Theorem 1, we write B^* as

$$\frac{1}{k_n}\sum_{j=1}^{k_n}\left[\left(X_j^*\right)^2 - (\bar{X})^2 - S_n^2\right].$$

Noting that $(X_j^*)^2 - (\bar{X})^2 - S_n^2$ $(j = 1, \ldots, n)$ are conditionally i.i.d. random variables and using the strong law of large numbers, we obtain $B^* \xrightarrow{a.s.} 0$ as $n \to \infty$ for almost all sample sequences of $X_1, X_2, \ldots$.

Finally, for any $\varepsilon > 0$, we obtain from the Markovian inequality that

$$P^*\left(\left|\frac{2X_i^*}{W_{in}}\right| \ge \varepsilon\right) \le \frac{1}{\varepsilon^2}E^*\left(\frac{2X_i^*}{W_{in}}\right)^2 = \frac{4}{W_{in}^2\varepsilon^2}E^*\left(X_i^*\right)^2 = \frac{4}{W_{in}^2\varepsilon^2}\frac{1}{n}\sum_{j=1}^{n} X_j^2.$$

Because $\frac{1}{n}\sum_{j=1}^{n} X_j^2 \xrightarrow{a.s.} \mu^2 + \sigma^2 < \infty$, we have, for almost all sample sequences of $X_1, X_2, \ldots$,

$$P^*\left(\left|\frac{2X_i^*}{W_{in}}\right| \ge \varepsilon\right) \to 0 \quad \text{as } n \to \infty.$$

That is, for almost all sample sequences of $X_1, X_2, \ldots$, it is true that

$$\frac{2X_i^*}{W_{in}} \xrightarrow{P^*} 0 \quad \text{as } n \to \infty,$$

which implies that $C^* \xrightarrow{P^*} 0$ as $n \to \infty$ for almost all sample sequences.

In conclusion, according to the Slutsky theorem and Lemma 6, we obtain $c_i^*(d) \rightsquigarrow c(d)$ as $n \to \infty$ for almost all sample sequences of $X_1, X_2, \ldots$. Equation (19) is then proved according to Lemma 1. □

Remark 4 In the proofs of the three theorems, different ways are used to prove the consistency of the bootstrap approximations for Cases 1 and 2. For Case 2, the distributions of

each local statistic and its bootstrap scenario are bridged by a same normal distribution. Therefore, it can be inferred that the bootstrap approximation performs at least as well as the normal distribution in this case. For Case 1, however, the numerator of each statistic is the sum of a fixed number of random variables in the process of $n \to \infty$. The limit distribution of each statistic cannot be a normal distribution if the population for drawing the sample does not follow a normal distribution. Therefore, the normal approximation fails to approximate the null distribution of each statistic in this case, but the bootstrap approximation still works according to the proof of each theorem, which is possibly the main reason for the empirical finding that the normal approximation is sometimes problematic as mentioned in the introduction. In practice, the neighbors of a reference location is generally very few relatively to the sample size and, as aforementioned, the bootstrap method can provide a valid approximation to the null distribution of each local statistic. In summary, the bootstrap approximation outperforms the normal approximation especially in practice.

4 Application to the spatial pattern detection of the Boston housing price data

In order to demonstrate the application of the bootstrap approximations, a real-world example based on the Boston housing price data is analyzed for the significance test of local spatial association. As mentioned in Remark 4, the bootstrap method can provide a valid approximation for the null distribution of each local statistic. However, for a local-statistic-based test with the bootstrap approximation, some other issues such as the Monte-Carlo implementation of the bootstrap method and the multiple test problem should be considered. The purpose of this section is to provide a full process of using the bootstrap approximation in practice.

4.1 Description of the data set and determination of the spatial linkage matrix

The Boston housing price data set, which is publicly available in the R package *spdep* (http://eran.r-project.org/), consists of observations of the median house value (in $1000) of owner-occupied homes and 13 explanatory variables in 506 US census tracts of the Boston area in 1970. Moreover, a list of influential neighbors for each tract is also attached, where a tract is an influential neighbor of another tract if these two tracts share a common part of the boundary.

Here, we chose the median house value, which we denoted by X henceforth, as the target variable to detect its spatial variation patterns based on the observations $x_1, x_2, \ldots, x_n$ of X in the $n = 506$ census tracts. The spatial linkage matrix $W = (w_{ij})_{n\times n}$ was obtained from the list of influential neighbors of each tract. Specifically, let $w_{ij} = 1$ if tract j is the influential neighbor of tract i; $w_{ij} = 0$ if otherwise; and $w_{ii} = 0$ by convention. The number of neighbors for the 506 census tract ranges from 1 to 8 with the averaged value being 4.25 which is much smaller than the sample size $n = 506$.

First of all, we conducted the Kolmogorov–Smirnov test for the normality of the observations of the target variable X. The p-value of the test is $p = 0.0000$, providing strong evidence of non-normality of the observations. As mentioned in Remark 4, the normal approximation to the null distributions of the three local statistics is problematic while the bootstrap approximation works for this data set.

4.2 Monte Carlo implementation of the bootstrap distribution functions

In general, the exact bootstrap distribution of a statistic is difficult to derive, although it is theoretically known for a given sample drawn from the population. In practice, Monte Carlo simulation is commonly used to compute the bootstrap distribution of the statistic. Here, we take the Getis and Ord's G_i statistic (we omit the distance threshold d in the statistic here because the spatial linkage matrix was determined without using it explicitly) as an example to show the Monte Carlo procedure. The procedure for the other two statistics is essentially the same.

Let $x_1, x_2, \ldots, x_n$ be the observations of the target variable X with x_i located at the location s_i. Given a reference location s_i, the Monte Carlo procedure for approximating the bootstrap distribution of G_i is as follows.

Step 1. Draw with replacement a bootstrap sample $(x_1^*, x_2^*, \ldots, x_n^*)$ from $(x_1, x_2, \ldots, x_n)$. Specifically, for each of $k = 1, 2, \ldots, n$, draw a random number u from the uniform distribution $U(0, 1)$, and let $x_k^* = x_{[nu]+1}$.

Step 2. Compute the bootstrap value G_i^* of G_i according to Eq. (4).

Step 3. Repeat Steps 1 and 2 for N times and obtain N bootstrap values of G_i which we denote $G_{i(1)}^*, G_{i(2)}^*, \ldots, G_{i(N)}^*$.

Step 4. Compute the empirical distribution function of $G_{i(1)}^*, G_{i(2)}^*, \ldots, G_{i(N)}^*$ and take it as an estimator of the bootstrap distribution of G_i. That is, for each real number x, the bootstrap distribution function of G_i is approximated by

$$P^*\left(G_i^* \le x\right) = \frac{1}{N}\sum_{k=1}^{N} \mathbb{I}_{\{G_{i(k)}^* \le x\}}. \tag{21}$$

4.3 Spatial association detection of the Boston housing price data

4.3.1 Alternative hypotheses and p-values of the tests

As pointed out by Getis and Ord [11], G_i measures the concentration or lack concentration of the values associated with the variable X on the reference location s_i. Therefore, G_i is commonly used to identify a location which is surrounded by large values or small values of X in its neighborhood. I_i and c_i can be employed to test whether the value of X located at the reference location s_i is similar (local positive autocorrelation) or dissimilar (local negative autocorrelation) to those located at its neighbors. To be specific, we mainly focused in this case study on identifying such a location that is surrounded by large values located at its neighbors by G_i, that is, a location with extremely large value of G_i, and testing local positive autocorrelation using I_i and c_i, that is, a location with extremely large value of I_i or with extremely small value of c_i. These above objectives amount to the G_i-, I_i- and c_i-based tests for the following alternative hypotheses, respectively:

H_{1G}: a tract surrounded by its neighbors with high housing price;

H_{1I}: a tract with the housing price being positively correlated to those in its neighbors;

H_{1c}: a tract with the housing price being similar to those in its neighbors.

The above alternative hypotheses all lead to one-sided tests. Specifically, the p-value of the G_i test derived by the bootstrap distribution in Eq. (21) is

$$p_{G_i} = \frac{1}{N}\sum_{k=1}^{N} \mathbb{I}_{\{G_{i(k)}^* \ge G_i^{(0)}\}}, \tag{22}$$

where $G_i^{(0)}$ is the observed value of G_i at the reference location s_i and is computed according to Eq. (1) with the sample $(X_1, X_2, \ldots, X_n)$ replaced by its observed value $(x_1, x_2, \ldots, x_n)$. Similarly, the p-values of the I_i test and the c_i test are, respectively,

$$p_{I_i} = \frac{1}{N} \sum_{k=1}^{N} \mathbb{I}_{\{I_{i(k)}^* \geq I_i^{(0)}\}} \tag{23}$$

and

$$p_{c_i} = \frac{1}{N} \sum_{k=1}^{N} \mathbb{I}_{\{c_{i(k)}^* \leq c_i^{(0)}\}}, \tag{24}$$

where $I_i^{(0)}$ and $c_i^{(0)}$ are the observed values of I_i and c_i computed according to Eqs. (2) and (3), respectively.

4.3.2 Method for dealing with multiple testing problem

When a local statistic is used to identify local spatial association of geo-referenced data, the test is generally performed at each location over the study region based on the same observations, which involves the multiple testing problem. Therefore, a given overall significance level, say α, should be properly adjusted in order to control the overall type I error to be less than α. Although the commonly used Bonferroni and Sidák criterions can readily be used here for adjusting the overall significance level, both methods are very conservative especially when the sample size is large [1]. Caldas and Singer [6] have used the so-called false discovery rate (FDR) criterion, developed by Benjamini and Hochberg [2], to handle the multiple testing problem associated with local spatial statistics and the results demonstrated that the FDR criterion is much more powerful than the Bonferroni and the Sidák methods. Therefore, the FDR criterion is employed here for dealing with the multiple testing problem in the analysis of the Boston housing price data with the G_i, I_i and c_i statistics. We introduce in what follows the FDR criterion in its general case.

Suppose that a total of K tests are simultaneously conducted based on a local statistic and the resultant p-values are $p_1, p_2, \ldots, p_K$, respectively. Sort the p-values in ascending order as $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(K)}$, and let

$$k_0 = \max\left\{k : p_{(k)} \leq \frac{k}{K}\alpha, k = 1, 2, \ldots, K\right\},$$

where α is the given overall significance level. The adjusted significance level for each individual test is $\alpha_A = \frac{k_0}{K}\alpha$.

4.3.3 Testing results with analysis

For the Boston housing price data, the sample size is $n = 506$. Given each of the three local statistics G_i, I_i and c_i, the bootstrap procedure was used to compute the p-value at each of the 506 tracts, in which the number of the bootstrap replications is $N = 500$. The overall significance level was set to be $\alpha = 0.05$. Using the FDR criterion, we saw that the adjusted significance levels are $\alpha_A^G = 0.00198$ for G_i, $\alpha_A^I = 0.00573$ for I_i, and $\alpha_A^c = 0.01107$ for c_i, respectively. The maps of the testing results are shown in Fig. 1, where the black areas represent the tracts with the original p-values being less than the overall significance level

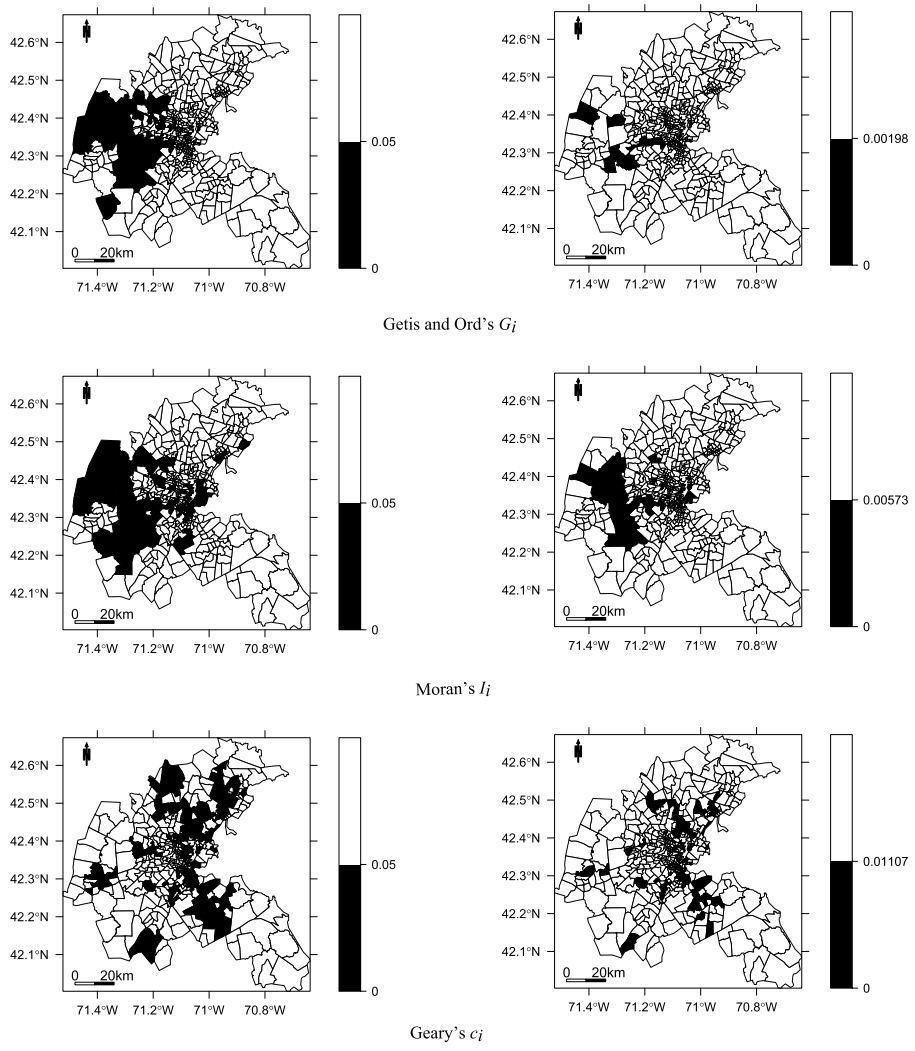

Figure 1 Maps of the categorized *p*-values for the G_i, I_i and c_i tests. The black areas represent the tracts with the *p*-values less than the overall significance level $\alpha = 0.05$ (left-hand column) or less than the corresponding adjusted significance levels $\alpha_A^G = 0.00198$, $\alpha_A^I = 0.00573$, and $\alpha_A^c = 0.01107$ (right-hand column)

$\alpha = 0.05$ (left-hand column) or less than the corresponding adjusted significance levels (right-hand column).

The result of the G_i test (panels in the first row) shows that the tracts with high housing price concentration appear mainly in the middle western region. After the adjustment of the overall significance level, only a few of tracts show the pattern that they are surrounded by their respective neighbors with high housing price.

The result of the I_i test (panels in the second row) shows a similar pattern to that of the G_i test especially under the overall significance level of $\alpha = 0.05$. That is, the tracts with similar housing price to those of their respective neighbors also locate on the middle western region except for some tracts on the middle eastern part. After the significance level is adjusted to $\alpha_A^I = 0.0057$, a belt region where positive spatial autocorrelation is significant is clearly shown. By the combination of the results from G_i and I_i tests, we know

that these common tracts colored in black and their respective neighbors all share high housing price, indicating "hot" spots of housing price in the Boston area.

The result of the c_i test (panels in the last row) demonstrates a totally opposite spatial pattern to that of the I_i test for the significant tracts, although both tests focus on detecting such tracts that share a similar housing price with their respective neighbors. Given the foregoing analysis showing that the I_i test uncovers the tracts with high housing price sharing with their respective neighbors, it can be inferred that the c_i test clarifies such tracts that share low house price with their respective neighbors. According to the structures of the I_i and c_i statistics, the opposite spatial patterns identified by the I_i and the c_i tests may imply that a large difference generally exists in the high housing price shared by a reference tract and its neighbors, while the low housing prices shared by a reference tract and its neighbors are relatively homogeneous. Moreover, it can be observed from the figure that the tracts sharing low housing price with their respective neighbors are more separately spatially distributed than those sharing high housing price with their respective neighbors. That is to say, the "cool" spots in the housing price are separately spatially distributed and the "hot" spots crowd in space.

5 Final remarks

There has been a growing interest in using local statistics to explore local patterns of spatial association in geo-referenced data, in which the null distributions of the local statistics play a key role in the related statistical inference. Considering that the bootstrap method can well account for non-normality of data and can easily be implemented with modern computers, we propose in this paper a bootstrap method to approximate the null distributions of the commonly used local spatial statistics of Getis and Ord's G_i, Moran's I_i and Geary's c_i. More importantly, strong consistency of the bootstrap approximation is established, which provides not only a theoretical basis for using the bootstrap method to approximate the null distributions of these three statistics, but also some evidence that normal approximation sometimes fails to approximate the null distributions of these local statistics. Furthermore, the practical implementation procedure of the local spatial statistics based bootstrap tests is fully given by a case study of the Boston housing price data.

Methodologically, the bootstrap procedure can readily be used to approximate the null distributions of other local spatial statistics such as Ord and Getis's LOSH statistic [19, 25]. However, establishing a common theoretical framework for the validity of the bootstrap approximation seems not easy. Therefore, consistency of the bootstrap approximation for other local spatial statistics or, furthermore, convergence rate of the current bootstrap approximation deserves to be investigated in the future research.

Acknowledgements
The authors would like to thank the reviewer for his/her valuable comments and suggestions, which led to significant improvement on the manuscript.

Funding
This work was supported by the National Nature Science Foundation of China (Nos. 11871056 and 11271296).

Availability of data and materials
The real-world data set is available in the R package "spdep" linked to http://eran.r-project.org/.

Competing interests
The authors declare no competing interests.

Authors' contributions
CLM contributed the idea, formulated the methodology and wrote part of the original draft; SFX completed the theoretical proofs and wrote part of the original draft. FC performed the computation of the real-word example. All authors read and approved the final manuscript.

Author details
[1]School of Science, Xi'an Polytechnic University, Xi'an, China. [2]School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China. [3]College of Mathematics and Information Science, Xinxiang University, Xinxiang, China.

Publisher's Note

Received: 15 January 2020 Accepted: 26 August 2020 Published online: 04 September 2020

References

1. Anselin, L.: Local indicators of spatial association—LISA. Geogr. Anal. **27**(2), 93–115 (1995)
2. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B **57**(1), 289–300 (1995)
3. Bickel, P.J., Freedman, D.A.: Some asymptotic theory for the bootstrap. Ann. Stat. **9**(6), 1196–1217 (1981)
4. Bivand, R., Müller, W.G., Reder, M.: Power calculations for global and local moran's *I*. Comput. Stat. Data Anal. **53**(8), 2859–2872 (2009)
5. Boots, B., Tiefelsdorf, M.: Global and local spatial autocorrelation in bounded regular tessellations. J. Geogr. Syst. **2**(4), 319–348 (2000)
6. Caldas, M.C., Singer, B.H.: Controlling the false discovery rate: a new application to account for multiple and dependent tests in local statistics of spatial association. Geogr. Anal. **38**(2), 180–208 (2006)
7. DasGupta, A.: Asymptotic Theory of Statistics and Probability. Springer, New York (2008)
8. Efron, B.: Bootstrap methods: another look at the jackknife. Ann. Stat. **7**(1), 1–26 (1979)
9. Getis, A.: Spatial filtering in a regression framework: examples using data on urban crime, regional inequality, and government expenditures. In: Anselin, L., Rey, S.J. (eds.) Perspectives on Spatial Data Analysis, pp. 191–202. Springer, Berlin (2010)
10. Getis, A., Griffith, D.A.: Comparative spatial filtering in regression analysis. Geogr. Anal. **34**(2), 130–140 (2002)
11. Getis, A., Ord, J.K.: The analysis of spatial association by use of distance statistics. Geogr. Anal. **24**(3), 189–206 (1992)
12. Hardisty, F., Klippel, A.: Analysing spatio-temporal autocorrelation with LISTA-Viz. Int. J. Geogr. Inf. Sci. **24**(10), 1515–1526 (2010)
13. Leung, Y., Mei, C.L., Zhang, W.X.: Statistical test for local patterns of spatial association. Environ. Plan. A **35**(4), 725–744 (2003)
14. Liu, X.Q., Sun, T.S., Li, G.P.: Spatial analysis of industry clusters based on local spatial statistics: a case study of Beijing manufacturing industry clusters. Sci. Geogr. Sin. **32**(5), 530–535 (2012)
15. Major, P.: On the invariance principle for sums of independent identically distributed random variables. J. Multivar. Anal. **8**(4), 487–517 (1978)
16. Mallows, C.L.: A note on asymptotic joint normality. Ann. Math. Stat. **43**(2), 508–515 (1972)
17. Mclaughlin, C.C., Boscoe, F.P.: Effects of randomization methods on statistical inference in disease cluster detection. Health Place **13**(1), 152–163 (2007)
18. Ord, J.K., Getis, A.: Local spatial autocorrelation statistics: distributional issues and an application. Geogr. Anal. **27**(4), 286–306 (1995)
19. Ord, J.K., Getis, A.: Local spatial heteroscedasticity (LOSH). Ann. Reg. Sci. **48**(2), 529–539 (2012)
20. Tiefelsdorf, M.: Some practical applications of Moran's *I*'s exact conditional distribution. Pap. Reg. Sci. **77**(2), 101–129 (1998)
21. Tiefelsdorf, M.: The saddlepoint approximation of Moran's *I*'s and local Moran's I_i's reference distributions and their numerical evaluation. Geogr. Anal. **34**(3), 187–206 (2002)
22. Tiefelsdorf, M., Boots, B.: The exact distribution of Moran's *I*. Environ. Plan. A **27**(6), 985–999 (1995)
23. van der Vaart, A.W.: Asymptotic Statistics. Cambridge University Press, New York (2000)
24. Xie, Z., Yan, J.: Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. J. Transp. Geogr. **31**(5), 64–71 (2013)
25. Xu, M., Mei, C.L., Yan, N.: A note on the null distribution of the local spatial heteroscedasticity (LOSH) statistic. Ann. Reg. Sci. **52**(3), 697–710 (2014)
26. Yan, N., Mei, C.L., Wang, N.: A unified bootstrap test for local patterns of spatiotemporal association. Environ. Plan. A **47**(1), 227–242 (2015)
27. Zhang, T.: Limiting distribution of the G statistics. Stat. Probab. Lett. **78**(12), 1656–1661 (2008)