CORRECTED MORAN'S I STATISTIC

by

Jian Ying

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

in

Computing

School of Computing

The University of Utah

March 2019

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Jian Ying**
has been approved by the following supervisory committee members:

Jeff M. Phillips ,	Chair(s)	**March 2019** Date Approved
Aditya Bhaskara ,	Member	**March 2019** Date Approved
Shandian Zhe ,	Member	**March 2019** Date Approved

by **Ross Whitaker** , Chair/Dean of
the Department/College/School of **Computing**
and by **David B. Kieda** , Dean of The Graduate School.

ABSTRACT

Moran's I is a statistic that measures spatial correlation of n spatial data points. It has been widely used in identifying spatial patterns. In this thesis, a method that estimates Moran's I by sampling and the bounds of error of the estimate is described. A corrected version of Moran's I that reduces bias of the estimate when the associated value can only measured with noise is proposed and the properties are proved analytically and demonstrated by simulation.

CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my committee chair Professor Jeff Phillips. Without his guidance and persistent help this dissertation would not have been possible. It took him a lot of extra effort to guide me, a part-time student without strong background in computer science, to finish the long journey. I really appreciate.

I would like to thank my committee members, Professor Aditya Bhaskara and Professor Shandian Zhe, for the useful comments, remarks and engagement through the learning process of this master thesis.

Finally, I want to thank my parents, Guangtai Ying and Huaizhi Song, and my sister Jing Ying, for their constant support. Special thanks to my wife, Minmin, for her love and care. Last but not least, to my son Kevin and my daughter Katherine, thank you for being good kids and you make my life much more fulfilled.

CHAPTER 1

INTRODUCTION

Moran's I is a statistic that measures spatial correlation of n spatial data points $P = \{p_1, p_2, \ldots, p_n\}$ [1], [2]. Each point p_i is associated with a value $x_i \in \mathbb{R}$. Let $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ be the mean of all x_i values, and let $\tilde{z}_i = x_i - \bar{x}$ be a centred value. Each pair of points $p_i, p_j \in P$ also have an associated similarity weight $w_{i,j} \in [0, 1]$; we often represent all of these pairs in a symmetric matrix W. For instance, $w_{i,j}$ may be 1 if two points p_i, p_j are within a fixed distance (or the spatial regions they represent are adjacent) and 0 otherwise. Or we may define $w_{i,j} = \exp(-\|p_i - p_j\|^2/2\sigma^2)$ for spatial points $p_i, p_j \in \mathbb{R}^d$. Then the Moran's I is defined as

$$I_{\text{true}} = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{z}_i w_{i,j} \tilde{z}_j}{\frac{1}{n}\sum_{i=1}^{n} \tilde{z}_i^2}. \tag{1.1}$$

Moran's I statistic tests whether there are some relationships between location and associated values. A statistically significant positive statistic indicates that the associated values of nearby locations are more clustered than randomly distributed. It have been widely used to identify the spatial pattern of ACS estimates, facilitating the analysis of regional convergence/inequality and corresponding policy and decision making. For example, it was used to study the relationship between lithium concentration in drinking water and suicide rate [3]. It was used in dialectology to measure the spatial variation of language [4]. It was also used to measure spatial genetic structure in plant populations[5]. Particularly, it is very useful in social study. For example, it was used to investigate criminal rate [6].

However, when the observation is uncertain, the value of the statistic and inference are no longer stable. No work to date has assessed the impacts of observation uncertainty on spatial autocorrelation identication, nor developed new spatial autocorrelation statistics that are robust to the input uncertainty. This research seeks to fill these gaps. Chapter 2 describes estimating Moran's I by sampling and the bounds of error of the estimate. Chapter 3 describes a corrected version of Moran's I that reduces bias of the estimate when

the value x_i can only measured with noise.

CHAPTER 2

ESTIMATE MORAN'S I BY SAMPLING

2.1 Understanding the Error from Sampling

In this section we discuss implications of error in the estimate of I_{true} when the data is generated by randomly sampling from a source. First, we identify the model variance implied by sampling and how this relates to the Gaussian distribution assumed above. Next, we show how to estimate I_{true} using only a random sample of pairs of data points; an important consideration for very large datasets. Finally, we show that this estimate can be improved by using importance sampling to select certain pairs with higher probability, then adjusting their weight in the final estimation to make it unbiased.

2.1.1 Modeling Data from Sampling

Consider the case where $S = \{p_1, p_2, \ldots, p_n\}$ is finite but large point set. The relationships measured by $w_{i,j}$ is fixed and easy to calculate, but the value x_i is hard to obtain. For example, x_i may be constructed as either an aggregate or average from a population associated with p_i. So it will be a saving of resources if we can estimate Moran's I with needed accuracy by sampling form S. Next we show that is doable.

First we define two sets $\mathbb{A} = \{a(i,j) = \tilde{z}_i\tilde{z}_j\}$ and $\mathbb{B} = \{b(i) = \tilde{z}_i^2\}$. Then we define two new random variables: A represents the distribution of random sample from set $\mathbb{A}$ and B represents the distribution of random sample from set $\mathbb{B}$. Then using $W = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}$, define

$$Z_w = \frac{1}{W}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{z}_i w_{i,j}\tilde{z}_j = \frac{1}{W}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}a(i,j),$$

$$Z_2 = \frac{1}{n}\sum_{i=1}^{n}\tilde{z}_i^2 = \frac{1}{n}\sum_{i=1}^{n} b_i.$$

We can see that Z_w is the weighted mean of the random variable A with weight given by $w_{i,j}$ and Z_2 is the mean of the random variable B. Also observe that $I_{\text{true}} = \dfrac{Z_w}{Z_2}$. Note that if all of the weights $w_{i,j}$ are known, then W is a known constant.

We can then estimate I_{true} in two steps. First, to estimate Z_2 we simply sample $\tilde{z}_i$s uniformly and apply Chebyshev's Inequality to get bounds. To estimate Z_w we can uniformly sample pairs points (i,j) and get the values of $\tilde{z}_i$ and $\tilde{z}_j$ as well as $w_{i,j}$. However, Z_w can also be estimated as weighted mean of $\tilde{z}_i\tilde{z}_j$ with weight given by $w_{i,j}$. We can improve performance of the estimation by importance sampling upon uniform sampling. Then we can apply Chebyshev's Inequality again to give bounds on this estimate.

2.1.2 Algorithm to estimate Z_w

Now we give an algorithm to estimate Z_w. Assuming that we have a fixed set of size n for which we want to calculate the Moran's I and that we have calculated the weight $w_{i,j}$ and $w_{i,\cdot} = \sum_{j=1}^{n}$. We denote the distribution with probability mass function (pmf) $P(X = i) = w_{i,\cdot}$ as F and the distribution with pmf $P(X = j) = w_{i,j}/w_{i,\cdot}$ as F_i. We want to sample n_1 pairs of points by importance sampling to estimate Z_w. The Algorithm 1, below, implements the importance sampling we want.

Algorithm 1: Importance sampling to estimate Z_w

```
set S = 0
for i ← 1 to n_1 do
    generate a random integer in i = [1, n] with a distribution F.
    generate a random integer in j = [1, n] with a distribution F_i.
    estimate z̃_i and z̃_j
    S = S + z̃_i z̃_j
return S/n_1
```

2.1.3 Importance Sampling of Pairs

We can improve upon the above result by applying importance sampling. We have the importance sampling property:

Property 2.1 Suppose we want to find $\mu = E(f(X)) = \int_D f(x)\, p(x)\, dx$ where p is a probability density function on $D \subseteq \mathbb{R}^d$ and f is the integrand. We take $p(x) = 0$ for all $x \notin D$. If q is a positive probability density function on $\mathbb{R}^d$, then $\mu = E_p(f(X)) = E_q(\frac{f(x)\, p(x)}{q(x)})$ where$E_p(.)$ denotes expectation for $X \sim p$ and $E_q(.)$ denotes expectation for $X \sim q$. When the distribution of X is discrete, simply change the integration to summation and density function to mass function and the theorem still holds.

Now we apply property 2.1 above to estimate Z_w.

We have $Z_w = \frac{1}{W}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{z}_i w_{i,j}\tilde{z}_j = \frac{n^2}{W}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{z}_i w_{i,j}\tilde{z}_j\frac{1}{n^2} \equiv \frac{n^2}{W}\mu$ where $\mu = \sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{z}_i w_{i,j}\tilde{z}_j\frac{1}{n^2}$ is the expectation of $f(i,j) \equiv \tilde{z}_i w_{i,j}\tilde{z}_j$, where $f(i,j) \sim p$ and $p(i,j) \equiv \frac{1}{n^2}$ is a mass function. Now let $q(i,j) = \dfrac{w_{i,j}}{W}$ and $q(i,j)$ is also a mass function. Then by theorem 2.1, we have $Z_w = \frac{n^2}{W}\mu = \frac{n^2}{W}E_q(\dfrac{f(i,j)p(i,j)}{q(i,j)}) = \frac{n^2}{W}E_q(\dfrac{(\tilde{z}_i w_{i,j}\tilde{z}_j)(\frac{1}{n^2})}{\frac{w_{i,j}}{W}}) = \frac{n^2}{W}E_q((\tilde{z}_i\tilde{z}_j)(\frac{W}{n^2}))) = E_q((\tilde{z}_i\tilde{z}_j))$.

Therefore the importance sampling estimate of Z_w is $\hat{Z_w} = \frac{1}{m}\sum_{i=1}^{m}(\tilde{z}_{i1}\tilde{z}_{i2})$, where $\tilde{z}_{i1}$ and $\tilde{z}_{i2}$ are the two values of the sampled $i'th$ pair and the probability of the $i'th$ pair to be sampled is $\dfrac{w_{i1,i2}}{W}$.

2.1.4 Approximation Bounds

Next, we prove that $\hat{I}$ estimated by sampling described above is close to I_{true} with high probability for large enough sample sizes. We will achieve this by proving that numerators $\hat{Z_w}$ and Z_w are close, then showing the denominators $\hat{Z_2}$ and Z_2 are also close. Then we will combine these facts together to provide an overall bound.

2.1.4.1 Chebyshev's Inequality

We first represent Chebyshev's Inequality that will be used in our proof.

- Chebyshev's Inequality: if X is an arbitrary random variable and $t > 0$, then

$$\mathsf{Pr}(|X - \mathsf{E}(X)| \geq t) \leq \frac{\mathsf{Var}(X)}{t^2}.$$

  Now consider a set $X_1, X_2, \ldots, X_n$ of n uncorrelated random variables (that is, $Cov(X_i, X_j) = 0$) and their sum $S_n = \sum_{i=1}^{n} X_i$. Since $Cov(X_i, X_j) = 0$, we have $\mathsf{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n}\mathsf{Var}(X_i)$. Then for any $t > 0$, Chebyshev's Inequality implies:

$$\mathsf{Pr}(|S_n - \mathsf{E}(S_n)| \geq t) \leq \frac{\mathsf{Var}(S_n)}{t^2} = \frac{\sum_{i=1}^{n}\mathsf{Var}(X_i)}{t^2}.$$

  Furthermore, let $\sigma^2 = \dfrac{1}{n}\sum_{i=1}^{n}\mathsf{Var}(X_i)$ then $\mathsf{Pr}(\frac{1}{n}|S_n - \mathsf{E}(S_n)| \geq d) \leq \frac{\sigma^2}{nd^2}$.

2.1.4.2 Closeness in $\hat{Z_w}$.

Lemma 2.1.1. *For any parameter $t > 0$, when $n > \frac{C_1}{t^2 d}$ then $\mathsf{Pr}[|\hat{Z_w} - Z_w| > tZ_w] \leq d$, where C_1 is a constant that is larger than $\mathsf{Var}(A)/{Z_w}^2$.*

Proof. We assume that $\mathsf{Var}(A) = V_A$, $mean(A) = \mu_A$. To estimate Z_w, we apply importance sampling to get a sample of size n from A. We have proved that $\hat{Z_w} = \frac{1}{n}\sum_{i=1}^{m}(A_i)$ is an unbiased estimate of Z_w. Also We have $\mathsf{Var}(\hat{Z_w}) = V_A/n$. Now we apply Chebyshev's Inequality to get bounds of estimates of $\hat{Z_w}$.

$$\mathsf{Pr}(|\hat{Z_w} - Z_w| \geq tZ_w) \leq \frac{\mathsf{Var}(\hat{Z_w})}{t^2 Z_w^2} = \frac{V_A}{nt^2 Z_w^2}.$$

So if $n > \frac{V_A}{dt^2 Z_w^2}$ then $\mathsf{Pr}(|\hat{Z_w} - Z_w| \geq tZ_w) \leq d$.
In a special case that we know that $V_A < C_1 Z_w^2$, then if $n > \frac{C_1}{dt^2}$, then $\mathsf{Pr}[|\hat{Z_w} - Z_w| > tZ_w] \leq d$.

□

2.1.4.3 Closeness in $\hat{Z_2}$.

Lemma 2.1.2. *For any parameter $t > 0$, when $n > \frac{C_2}{t^2 d}$ then $\mathsf{Pr}[|\hat{Z_2} - Z_2| > tZ_2] \leq d$, where C_2 is a constant that is larger than $\mathsf{Var}(B)/{Z_2}^2$.*

Proof. We assume that $\mathsf{Var}(B) = V_B$ and $mean(B) = \mu_B$. To estimate Z_2, we uniformly sample a sample of size n from B. Obviously $\hat{Z_2} = \frac{1}{n}\sum_{i=1}^{m}(B_i)$ is an unbiased estimate of Z_2.
Also We have $\mathsf{Var}(\hat{Z_2}) = V_B/n$ Now we apply Chebyshev's Inequality to get bounds of estimates of

$$\mathsf{Pr}(|\hat{Z_2} - Z_2| \geq tZ_2) \leq \frac{\mathsf{Var}(\hat{Z_2})}{t^2 Z_2^2} = \frac{V_B}{nt^2 Z_2^2}.$$

So if $n > \frac{V_B}{dt^2 Z_2^2}$ then $\mathsf{Pr}(|\hat{Z_2} - Z_2| \geq tZ_2) \leq d$. In a special case that we know that $V_B < C_2 Z_2^2$, then if $n > \frac{C_2}{dt^2}$ then $\mathsf{Pr}[|\hat{Z_2} - Z_2| > tZ_2] \leq d$. □

2.1.4.4 Putting them together

Finally we put these results about the numerator and denominator stability together.

Theorem 2.1.1. *Consider a dataset P of size m where each value point has an associated value z_i. Define a random variable $A(i) = z_{i_1} z_{i_2}$ where z_{i_1} and z_{i_2} are a pair of values and $B(i) = z_{i_1}^2$. If we apply importance sampling with $q(i) = w(i) = w_{i_1,i_2}$to sample n_1 pairs of points to estimate Z_w and uniformly obtain another sample of size n_2 to estimate Z_2, Then, for any parameter $\alpha \in (0,1)$, if $n_1 > \frac{18C_1}{\delta\alpha^2}$ and $n_2 > \frac{18C_2}{\delta\alpha^2}$, with probability at least $1-\delta$, we have*

$$1-\alpha \leq \frac{\hat{I}}{I_{true}} \leq 1+\alpha.$$

Proof. Overall, when all the conditions mentioned in the proof above hold, which are $n_1 > \frac{C_1}{dt^2}$ and $n_2 > \frac{C_2}{dt^2}$, then

$$\mathsf{Pr}(\hat{Z}_2/Z_2 \leq 1-t \text{ or } \hat{Z}_2/Z_2 \geq 1+t) = \mathsf{Pr}(|\hat{Z}_2 - Z_2| \geq tZ_2) \leq d$$

and

$$\mathsf{Pr}(\hat{Z_w}/Z_w \leq 1-t \text{ or } \hat{Z_w}/Z_w \geq 1+t) = \mathsf{Pr}(|\hat{Z_w} - Z_w| \geq tZ_w) \leq d$$

hold and then

$$\mathsf{Pr}((\hat{Z_w}/Z_w)/(\hat{Z}_2/Z_2) < (1-t)/(1+t) \text{ or } (\hat{Z_w}/Z_w)/(\hat{Z}_2/Z_2) < (1+t)/(1-t)) \leq (2d).$$

And since $(\hat{Z_w}/Z_w)/(\hat{Z}_2/Z_2) = \hat{I}/I_{\text{true}}$ so

$$\mathsf{Pr}(\hat{I}/I_{\text{true}} < (1-t)/(1+t) \text{ or } \hat{I}/I_{\text{true}} > (1+t)/(1-t)) \leq (2d).$$

So

$$\mathsf{Pr}((1-t)/(1+t) \leq \hat{I}/I_{\text{true}} \leq (1+t)/(1-t)) \geq (1-2d).$$

Now, we rewrite the conditions and inequality in terms of α and δ.
Let $t = \frac{\alpha}{2+\alpha}$ then

$$\begin{aligned}(1+t)/(1-t) &= (1+\frac{\alpha}{2+\alpha})/(1-\frac{\alpha}{2+\alpha}) \\ &= (2+\alpha+\alpha)/(2+\alpha-\alpha) \\ &= (2+2\alpha)/2 \\ &= (1+\alpha)\end{aligned}$$

and

$$\begin{aligned}(1-t)/(1+t) &= (1-\frac{\alpha}{2+\alpha})/(1+\frac{\alpha}{2+\alpha}) \\ &= (2+\alpha-\alpha)/(2+\alpha+\alpha) \\ &= 2/(2+2\alpha) \\ &= 1/(1+\alpha) \\ &= 1-\alpha/(1+\alpha) \\ &> 1-\alpha\end{aligned}$$

Also let $d = \frac{1}{2}\delta$ then

$$1 - 2d = 1 - \delta$$

So

$$\Pr(1 - \alpha \leq I_{\text{corr}} / I_{\text{true}} \leq 1 + \alpha) \geq (1 - \delta).$$

Now plug $t = \dfrac{\alpha}{2+\alpha}$ and $\varepsilon = \frac{1}{2}\delta$ in to the conditions we mentioned in the proof of Closeness in denominators and denominators. We get the conditions in terms of α and δ and they are

- $n_1 > \frac{C_1}{(\frac{\alpha}{2+\alpha})^2 \frac{1}{2}\delta} = \frac{2(2+\alpha)^2 C_1}{\alpha^2 \delta}$
- $n_2 > \frac{C_2}{(\frac{\alpha}{2+\alpha})^2 \frac{1}{2}\delta} = \frac{2(2+\alpha)^2 C_2}{\alpha^2 \delta}$

Since $0 < \alpha < 1$, conditions below is looser than conditions above.

- $n_1 > \frac{18 C_1}{\alpha^2 \delta}$
- $n_2 > \frac{18 C_2}{\alpha^2 \delta}$

That finishes the proof of the theorem.

□

2.1.5 Simulation

A simulation study was performed to investigate how sample size of the sampling algorithm affect the accuracy of Moran's I estimation. The true attribute value x_i for each point p_i on an 50 by 50 grid is generated using R package *gstat*. The weight matrix W is defined so $w_{i,j}$ is 1 when p_i and p_j are adjacent on the grid, and 0 otherwise. True Moran's I was calculated using formula 1.1 as the reference. The sampling algorithms was applied to the generated data for 100 times for each of tested sample sizes. The Moran's I estimation for different sample sizes are summarized by box plot on figure 2.1. It is clearly shown that with sample size increases, the estimation of I approaches the true I.

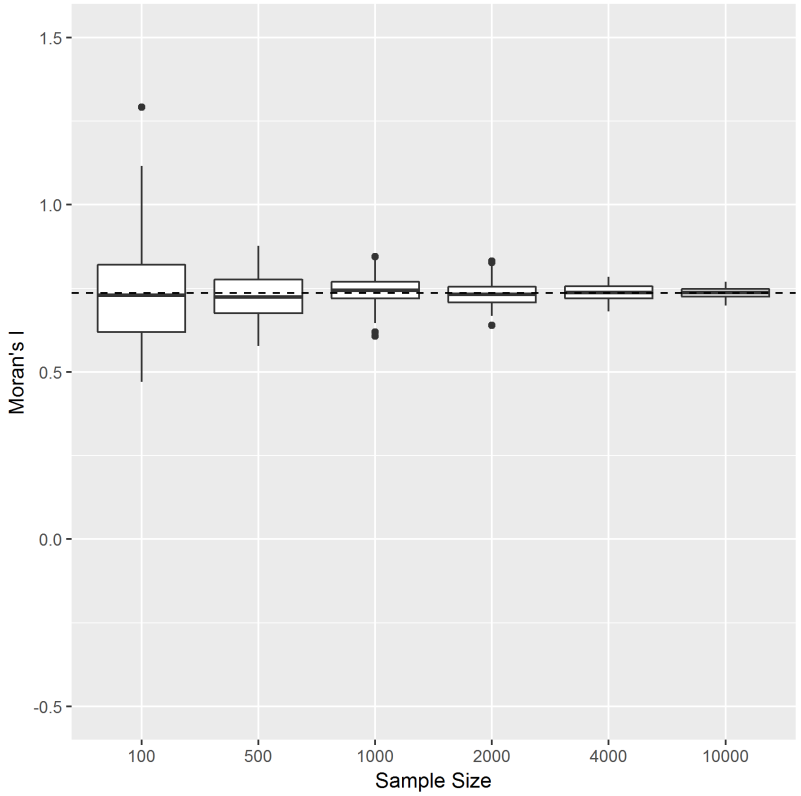

Figure 2.1. The effect of sample size on accuracy of Moran's I estimation

CHAPTER 3

CORRECTED MORAN'S I

3.1 Corrected Moran's I when there is noise

Very often, the true value associated with point p_i (denoted $\tilde{z}_i$) is not observed. Instead what we observe is $z_i = \tilde{z}_i + \varepsilon_i$, where ε_i is an unknown noise term, and we assume $\varepsilon_i \perp \varepsilon_j$ (they are independent). We will also assume that $\varepsilon_i \sim N(0, v_i)$. It is now typically for uncertain data to provide some information from which we can at least construct a good estimate of this. When z_i (instead of $\tilde{z}_i$) is used to calculate Moran's I, the estimated Moran's I will be smaller than the true one. This is similar to what happens in the calculation of the correlation coefficient. Let $\beta = \tilde{\beta} + \varepsilon_\beta$ and $\theta = \tilde{\theta} + \varepsilon_\theta$ where $\tilde{\beta}$ and $\tilde{\theta}$ are the true values and β and θ are estimate of $\tilde{\beta}$ and $\tilde{\theta}$, respectively, where ε_β and ε_θ are the associated measurement errors. Then the observed correlation $\mathsf{corr}(\beta, \theta)$ is smaller than the true correlation $\mathsf{corr}(\tilde{\beta}, \tilde{\theta})$. This phenomenon is called *disattenuation* and can be corrected by the equation: $\mathsf{corr}(\tilde{\beta}, \tilde{\theta}) = \mathsf{corr}(\beta, \theta)/\sqrt{R_\beta R_\theta}$ where $R_\beta = (\mathsf{Var}(\beta) - \mathsf{Var}(\varepsilon_\beta))/\mathsf{Var}(\beta)$ and $R_\theta = (\mathsf{Var}(\theta) - \mathsf{Var}(\varepsilon_\theta))/\mathsf{Var}(\theta)$.

Inspired by the correction of disattenuation in correlation coefficient, we propose to correct the Moran's I by

$$I_{\text{corr}} = \frac{1}{\sum_i \sum_j \omega_{i,j}} \frac{\sum_i \sum_j z_i \omega_{i,j} z_j}{\frac{1}{n}\sum_i z_i^2 - \sum_i v_i}. \tag{3.1}$$

3.1.1 Approximation Bounds

Next, we prove that I_{corr} is close to I_{true} with high probability for large enough dataset sizes. We will achieve this by proving that numerators $N \equiv \sum_i \sum_j z_i \omega_{i,j} z_j$ and $N_0 \equiv \sum_i \sum_j \tilde{x}_i \omega_{i,j} \tilde{x}_j$ are close, then showing the denominators $D \equiv \frac{1}{n}\sum_i z_i^2 - \sum_i v_i$ and $D_0 \equiv \frac{1}{n}\sum_i \tilde{z}_i^2$ are also close. Then we will combine these facts together to provide an overall

bound. We assume that the noise is bounded such that $v_i \leq v = c_1 D_0$ (c_1 *is a constant*). By definition, D_0 is the variance of the true value among the locations, and v is the bound of noise. So c_1 can be seem as the ratio of noise to signal. Since rescaling the weight w_{ij} doesn't change I_{corr}, without losing generality, we can assume $0 \leq \omega_{i,j} \leq 1$, and in this paper we only consider the case that the weight matrix is sparse such that $\sum_i \omega_{i,j} = \sum_j \omega_{i,j} < w \, \forall i$ but not too sparse such that $W = \sum_i \sum_j \omega_{i,j} > c_2 n$.

We first list a few standard properties about random variables X and Y we will use.

(P1) $\mathsf{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y)$

(P2) If $X \perp Y$ then $\mathsf{E}(XY) = \mathsf{E}(X)\mathsf{E}(Y)$.
that is, if X and Y are independent, then we can decompose the product of their expected values.

(P3) If $X \perp Y$ then $\mathsf{Var}(XY) = \mathsf{E}(X)^2\mathsf{Var}(Y) + \mathsf{E}(Y)^2\mathsf{Var}(X) + \mathsf{Var}(X)\mathsf{Var}(Y)$, and further if $\mathsf{E}(X) = 0$ and $\mathsf{E}(Y) = 0$ then $\mathsf{Var}(XY) = \mathsf{Var}(X)\mathsf{Var}(Y)$.

(P4) Chebyshev's Inequality: if X is an arbitrary random variable and $t > 0$, then

$$\mathsf{Pr}(|X - \mathsf{E}(X)| \geq t) \leq \frac{\mathsf{Var}(X)}{t^2}.$$

Now consider a set $X_1, X_2, \ldots, X_n$ of n uncorrelated random variables (that is,$Cov(X_i, X_j) = 0$) and their sum $S_n = \sum_{i=1}^n X_i$. Since $Cov(X_i, X_j) = 0$, we have $\mathsf{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathsf{Var}(X_i)$. Then for any $t > 0$, Chebyshev's Inequality implies:

$$\mathsf{Pr}(|S_n - \mathsf{E}(S_n)| \geq t) \leq \frac{\mathsf{Var}(S_n)}{t^2} = \frac{\sum_{i=1}^n \mathsf{Var}(X_i)}{t^2}.$$

Furthermore, let $\sigma^2 = \frac{1}{n}\sum_{i=1}^n \mathsf{Var}(X_i)$ then $\mathsf{Pr}(\frac{1}{n}|S_n - \mathsf{E}(S_n)| \geq d) \leq \frac{\sigma^2}{nd^2}$.

3.1.1.1 Closeness in numerators.

Lemma 3.1.1. *For any parameter $t > 0$, when $n > \frac{w(2+c_1)c_1}{t^2(I_{true}c_2)^2 d}$ then $\mathsf{Pr}[|N - N_0| > tN_0] \leq d$*

Proof. We first expand the definition of N

$$N = \sum_i \sum_j z_i \omega_{i,j} z_j = \sum_i \sum_j (\tilde{z}_i + \varepsilon_i)\omega_{i,j}(\tilde{z}_j + \varepsilon_j) = \sum_i \sum_j (\tilde{z}_i\tilde{z}_j + \tilde{z}_i\varepsilon_j + \tilde{z}_j\varepsilon_i + \varepsilon_i\varepsilon_j)w_{ij}.$$

Since $\mathsf{E}[\varepsilon_i] = 0$ and $\varepsilon_i \perp \varepsilon_j$, then $\mathsf{E}(\tilde{z}_i\varepsilon_j) = \mathsf{E}(\tilde{z}_j\varepsilon_i) = \mathsf{E}(\varepsilon_i\varepsilon_j) = 0$, hence $\mathsf{E}(N) = \mathsf{E}(N_0)$.

Next we can argue that ε_i and $\varepsilon_i\varepsilon_j$ are uncorrelated since

$$\mathsf{Cov}(\varepsilon_i, \varepsilon_i\varepsilon_j) = \mathsf{E}(\varepsilon_i\varepsilon_i\varepsilon_j) - \mathsf{E}(\varepsilon_i)\mathsf{E}(\varepsilon_i\varepsilon_j) = \mathsf{E}(\varepsilon_i^2)\mathsf{E}(\varepsilon_j) - \mathsf{E}(\varepsilon_i)\mathsf{E}(\varepsilon_i\varepsilon_j) = 0 - 0 = 0.$$

Moreover we have $\varepsilon_i\varepsilon_k$ and $\varepsilon_\ell\varepsilon_j$ are uncorrelated for any $i \neq k \neq j \neq \ell$ because

$$\mathsf{Cov}(\varepsilon_i\varepsilon_k, \varepsilon_i\varepsilon_j) = \mathsf{E}(\varepsilon_i\varepsilon_k\varepsilon_i\varepsilon_j) - \mathsf{E}(\varepsilon_i\varepsilon_k)\mathsf{E}(\varepsilon_i\varepsilon_j) = \mathsf{E}(\varepsilon_i^2)\mathsf{E}(\varepsilon_k)\mathsf{E}(\varepsilon_j) - \mathsf{E}(\varepsilon_i)\mathsf{E}(\varepsilon_k)\mathsf{E}(\varepsilon_i\varepsilon_j) = 0 - 0 = 0.$$

So all the terms in N are uncorrelated from each other. Also $\mathsf{Var}(\tilde{z}_i\varepsilon_j) = \tilde{z}_i^2 v_i$ and $\mathsf{Var}(\varepsilon_i\varepsilon_j) = \mathsf{Var}(\varepsilon_i)\mathsf{Var}(\varepsilon_j) = v_i v_j$ due to that $\varepsilon_i \perp \varepsilon_j$ and $\mathsf{E}(\varepsilon_i) = \mathsf{E}(\varepsilon_j) = 0$.So $Var(N) = \sum_i \sum_j (\tilde{z}_i^2 v_j + \tilde{z}_j^2 v_i + v_i v_j)\omega_{ij}^2$ According to Chebyshev's Inequality,

$$\mathsf{Pr}(|N - N_0| > tN_0) \leq \frac{\sum_i \sum_j (\tilde{z}_i^2 v_j + \tilde{z}_j^2 v_i + v_i v_j)\omega_{ij}^2}{t^2 N_0^2}.$$

In a special case where $v_i \leq v = c_1 D_0$ (c_1 is a constant), $0 \leq \omega_{i,j} \leq 1$, and the weight matrix is sparse such that $\sum_i \omega_{i,j} = \sum_j \omega_{i,j} < w \, \forall i$ but not too sparse such that $\sum_i \sum_j \omega_{i,j} = W > c_2 n$ then

$$\begin{aligned}
\mathsf{Pr}(|N - N_0| > tN_0) &\leq \frac{\sum_i \sum_j (\tilde{z}_i^2 v_j + \tilde{z}_j^2 v_i + v_i v_j)\omega_{ij}^2}{t^2 N_0^2} \leq \frac{\sum_i \sum_j (\tilde{z}_i^2 + \tilde{z}_j^2 + v) v \omega_{ij}}{t^2 N_0^2} \\
&= \frac{(\sum_i \sum_j \tilde{z}_i^2 \omega_{ij} + \sum_i \sum_j \tilde{z}_j^2 \omega_{ij} + \sum_i \sum_j v\omega_{ij})v}{t^2 N_0^2} \\
&\leq \frac{(\sum_i \tilde{z}_i^2 w + \sum_j \tilde{z}_j^2 w + \sum_i vw)v}{t^2 N_0^2} \\
&= \frac{(nD_0 w + nD_0 w + nvw)v}{t^2 N_0^2} \\
&= \frac{(nD_0 w + nD_0 w + nc_1 D_0 w)c_1 D_0}{t^2 N_0^2} \\
&= \frac{nD_0^2 w(2 + c_1)c_1}{t^2 N_0^2} \\
&= \frac{nD_0^2 w(2 + c_1)c_1}{t^2 (I_{true} D_0 W)^2} \\
&= \frac{nw(2 + c_1)c_1}{t^2 (I_{true} W)^2} \\
&\leq \frac{nw(2 + c_1)c_1}{t^2 (I_{true} c_2 n)^2} \\
&= \frac{nw(2 + c_1)c_1}{t^2 (I_{true} c_2 n)^2} \\
&= \frac{w(2 + c_1)c_1}{t^2 (I_{true} c_2)^2 n} \\
&\equiv p_1
\end{aligned}$$

□

In the case that n is a big number, p_1 is a small number. Specifically, when $n > \frac{w(2+c_1)c_1}{t^2(I_{true}c_2)^2 d}$, then $p_1 < d$

3.1.1.2 Closeness in denominators.

Lemma 3.1.2. *For any parameter $t > 0$, when $n > \frac{(4+2c_1)c_1}{dt^2}$ then $\mathsf{Pr}(|D - D_0| > tD_0) < d$*

Proof. Recall

$$D \equiv \frac{1}{n}\sum_i z_i^2 - \sum_i v_i = \frac{1}{n}\sum_i((\tilde{z}_i + \varepsilon_i)^2 - v_i) = \frac{1}{n}\sum_i(\tilde{z}_i^2 + 2\tilde{z}_i\varepsilon_i + \varepsilon_i^2 - v_i).$$

Moreover, since $\mathsf{E}(\tilde{z}_i\varepsilon_i) = 0$ and $\mathsf{Var}(\tilde{z}_i\varepsilon_i) = \tilde{z}_i^2 v_i$ and because $\varepsilon_i \sim N(0, v_i)$ then hence $\frac{\varepsilon_i^2}{v_i} \sim \chi_1^2$ and $\mathsf{E}(\frac{\varepsilon_i^2}{v_i}) = 1$. This implies $\mathsf{E}(\varepsilon_i^2 - v_i) = 0$, that $\mathsf{Var}(\frac{\varepsilon_i^2}{v_i}) = 2$ and $\mathsf{Var}(\varepsilon_i) = 2v_i^2$.

Also

$$\mathsf{Var}(2\tilde{z}_i\varepsilon_i + \varepsilon_i^2) = \mathsf{Var}(2\tilde{z}_i\varepsilon_i) + 2\mathsf{Cov}(2\tilde{z}_i\varepsilon_i, \varepsilon_i^2) + \mathsf{Var}(\varepsilon_i^2)$$

But

$$\mathsf{Cov}(2\tilde{z}_i\varepsilon_i, \varepsilon_i^2) = \mathsf{E}(2\tilde{z}_i\varepsilon_i\varepsilon_i^2) - \mathsf{E}(2\tilde{z}_i\varepsilon_i)\mathsf{E}(\varepsilon_i^2) = 2\tilde{z}_i\mathsf{E}(\varepsilon_i^3) - 2\tilde{z}_i\mathsf{E}(\varepsilon_i)\mathsf{E}(\varepsilon_i^2) = 2\tilde{z}_i 0 - 2\tilde{z}_i 0\mathsf{E}(\varepsilon_i^2) = 0$$

The reason that $\mathsf{E}(\varepsilon_i^3) = 0$ is that $\varepsilon_i \sim N(0, V_i)$ which is a distribution symmetric about 0. So

$$\mathsf{Var}(2\tilde{z}_i\varepsilon_i + \varepsilon_i^2) = \mathsf{Var}(2\tilde{z}_i\varepsilon_i) + 2\mathsf{Cov}(2\tilde{z}_i\varepsilon_i, \varepsilon_i^2) + \mathsf{Var}(\varepsilon_i^2) = 4\tilde{z}_i^2 v_i + 2v_i^2$$

and $\mathsf{E}(D) = \mathsf{E}(D_0)$ due to that $\mathsf{E}(\tilde{z}_i\varepsilon_i) = 0$ and $\mathsf{E}(\varepsilon_i^2 - V_i) = 0$.

In summary, $\mathsf{E}(D) = \mathsf{E}(D_0)$ and $Var(D) = \frac{1}{n^2}\sum_i(4\tilde{z}_i^2 v_i + 2v_i^2)$. Now according to Chebyshev's Inequality,

$$\begin{aligned}
\mathsf{Pr}(|D - D_0| > tD_0) &\leq \frac{\frac{1}{n^2}\sum_i (4\tilde{z}_i^2 v_i + 2v_i^2)}{t^2 D_0^2} \\
&= \frac{\sum_i (4\tilde{z}_i^2 v_i + 2v_i^2)}{n^2 t^2 D_0^2} \\
&\leq \frac{\sum_i (4\tilde{z}_i^2 + 2v)v}{n^2 t^2 D_0^2} \\
&= \frac{(4nD_0 + 2nv)v}{n^2 t^2 D_0^2} \\
&= \frac{(4nD_0 + 2nc_1 D_0)c_1 D_0}{n^2 t^2 D_0^2} \\
&= \frac{(4 + 2c_1)c_1}{nt^2} \\
&\equiv p_2
\end{aligned}$$

and when n is a big number p_2 is small. In particular, when $n > \frac{(4+2c_1)c_1}{dt^2}$ then $p_2 < d$. □

3.1.1.3 Putting these together.

Finally we put these results about the numerator and denominator stability together.

Theorem 3.1.1. *Consider a dataset P of size n where each value $z_i = \tilde{z}_i + \varepsilon_i$ has independent Gaussian noise $\varepsilon_i \sim \mathcal{N}(0, v_i)$, that $v_i \leq v = c_1 D_0$ (c_1 is a constant), $0 \leq \omega_{i,j} \leq 1$, and the weight matrix is sparse such that $\sum_i \omega_{i,j} = \sum_j \omega_{i,j} < w \,\forall i$ but not too sparse such that $W = \sum_i \sum_j \omega_{i,j} > c_2 n$. Then, for any parameter $\alpha \in (0,1)$, if $n > max(18w \cdot \frac{(2+c_1)c_1}{(c_2)^2} \cdot \frac{1}{(I_{true})^2} \cdot \frac{1}{\alpha^2} \cdot \frac{1}{\delta}, 36(2+c_1)c_1 \cdot \frac{1}{\alpha^2} \cdot \frac{1}{\delta})$, with probability at least $1 - \delta$, we have*

$$1 - \alpha \leq \frac{I_{corr}}{I_{true}} \leq 1 + \alpha.$$

Proof. Overall, when all the conditions mentioned in the proof above hold, which are $n > \frac{w(2+c_1)c_1}{t^2(I_{true}c_2)^2 d}$ and $n > \frac{(4+2c_1)c_1}{dt^2}$, then

$$\mathsf{Pr}(D/D_0 \leq 1 - t \;\text{ or }\; D/D_0 \geq 1 + t) = \mathsf{Pr}(|D - D_0| \geq tD_0) \leq d$$

and

$$\mathsf{Pr}(N/N_0 \leq 1 - t \;\text{ or }\; N/N_0 \geq 1 + t) = \mathsf{Pr}(|N - N_0| \geq tN_0) \leq d$$

hold and then

$$\mathsf{Pr}((N/N_0)/(D/D_0) < (1-t)/(1+t) \;\text{ or }\; (N/N_0)/(D/D_0) < (1+t)/(1-t)) \leq (2d).$$

And since $(N/N_0)/(D/D_0) = I_{\text{corr}}/I_{\text{true}}$ so

$$\mathsf{Pr}(I_{\text{corr}}/I_{\text{true}} < (1-t)/(1+t) \text{ or } I_{\text{corr}}/I_{\text{true}} > (1+t)/(1-t)) \leq (2d).$$

So

$$\mathsf{Pr}((1-t)/(1+t) \leq I_{\text{corr}}/I_{\text{true}} \leq (1+t)/(1-t)) \geq (1-2d).$$

Now, we rewrite the conditions and inequality in terms of α and δ.
Let $t = \dfrac{\alpha}{2+\alpha}$ then

$$\begin{aligned}(1+t)/(1-t) &= (1+\frac{\alpha}{2+\alpha})/(1-\frac{\alpha}{2+\alpha}) \\ &= (2+\alpha+\alpha)/(2+\alpha-\alpha) \\ &= (2+2\alpha)/2 \\ &= (1+\alpha)\end{aligned}$$

and

$$\begin{aligned}(1-t)/(1+t) &= (1-\frac{\alpha}{2+\alpha})/(1+\frac{\alpha}{2+\alpha}) \\ &= (2+\alpha-\alpha)/(2+\alpha+\alpha) \\ &= 2/(2+2\alpha) \\ &= 1/(1+\alpha) \\ &= 1-\alpha/(1+\alpha) \\ &> 1-\alpha\end{aligned}$$

Also let $d = \frac{1}{2}\delta$ then

$$1-2d = 1-\delta$$

So

$$\mathsf{Pr}(1-\alpha \leq I_{\text{corr}}/I_{\text{true}} \leq 1+\alpha) \geq (1-\delta).$$

Now plug $t = \dfrac{\alpha}{2+\alpha}$ and $d = \frac{1}{2}\delta$ in to the conditions we mentioned in the proof of Closeness in denominators and denominators. We get the conditions in terms of α and δ and they are

- $n > \frac{w(2+c_1)c_1}{(\frac{\alpha}{2+\alpha})^2(I_{true}c_2)^2\frac{1}{2}\delta} = \frac{2w(2+\alpha)^2(2+c_1)c_1}{\alpha^2(I_{true}c_2)^2\delta}$

- $n > \frac{2(2+c_1)c_1}{(\frac{\alpha}{2+\alpha})^2 \frac{1}{2}\delta} = \frac{4(2+\alpha)^2(2+c_1)c_1}{\alpha^2\delta}$

Since $0 < \alpha < 1$,conditions below is looses than conditions above.

- $n > \frac{18w(2+c_1)c_1}{\alpha^2(I_{true}c_2)^2\delta} = 18w \cdot \frac{(2+c_1)c_1}{(c_2)^2} \cdot \frac{1}{(I_{true})^2} \cdot \frac{1}{\alpha^2} \cdot \frac{1}{\delta}$
- $n > \frac{36(2+c_1)c_1}{\alpha^2\delta} = 36(2+c_1)c_1 \cdot \frac{1}{\alpha^2} \cdot \frac{1}{\delta}$

That finishes the proof of the theorem.

□

Now we convert the relative bound to additive bound.

Theorem 3.1.2. *Consider a dataset P of size n where each value $z_i = \tilde{z}_i + \varepsilon_i$ has independent Gaussian noise $\varepsilon_i \sim \mathcal{N}(0, v_i)$, that $v_i \leq v = c_1 D_0 (c_1 \text{ is a constant}), 0 \leq \omega_{i,j} \leq 1$, and the weight matrix is sparse such that $\sum_i \omega_{i,j} = \sum_j \omega_{i,j} < w \, \forall i$ but not too sparse such that $W = \sum_i \sum_j \omega_{i,j} > c_2 n$. Then, for any parameter $\beta \in (0,1)$, if $n > max(18w \cdot \frac{(2+c_1)c_1}{(c_2)^2} \cdot \frac{1}{\beta^2} \cdot \frac{1}{\delta}, 36(2+c_1)c_1 \cdot \frac{1}{\beta^2} \cdot \frac{1}{\delta})$, with probability at least $1-\delta$, we have*

$$I_{true} - \beta \leq I_{corr} \leq I_{true} + \beta.$$

Proof. According to theorem 1.1, when $n > max(18w \cdot \frac{(2+c_1)c_1}{(c_2)^2} \cdot \frac{1}{(I_{true})^2} \cdot \frac{1}{\alpha^2} \cdot \frac{1}{\delta}, 36(2+c_1)c_1 \cdot \frac{1}{\alpha^2} \cdot \frac{1}{\delta})$, with probability at least $1-\delta$, we have

(E1)

$$1 - \alpha \leq \frac{I_{\text{corr}}}{I_{\text{true}}} \leq 1 + \alpha.$$

That is equivalent to

(E2)

$$I_{\text{true}} - \alpha I_{\text{true}} \leq I_{\text{corr}} \leq I_{\text{true}} + \alpha I_{\text{true}}.$$

Now let $\beta = \alpha I_{\text{true}}$ then we have $\alpha = \frac{\beta}{I_{\text{true}}}$ and E2 becomes

(E3)

$$I_{\text{true}} - \beta \leq I_{\text{corr}} \leq I_{\text{true}} + \beta.$$

Plug $\beta = \alpha I_{\text{true}}$ and $\alpha = \frac{\beta}{I_{\text{true}}}$ in $n > max(18w \cdot \frac{(2+c_1)c_1}{(c_2)^2} \cdot \frac{1}{(I_{true})^2} \cdot \frac{1}{\alpha^2} \cdot \frac{1}{\delta}, 36(2+c_1)c_1 \cdot \frac{1}{\alpha^2} \cdot \frac{1}{\delta})$ and because $-1 \leq I_{\text{true}} \leq 1$, we can easily get that $n > max(18w \cdot \frac{(2+c_1)c_1}{(c_2)^2} \cdot \frac{1}{\beta^2} \cdot \frac{1}{\delta}, 36(2+$

$c_1)c_1 \cdot \frac{1}{\beta^2} \cdot \frac{1}{\delta})$ is tighter than $n > max(18w \cdot \frac{(2+c_1)c_1}{(c_2)^2} \cdot \frac{1}{(I_{true})^2} \cdot \frac{1}{\alpha^2} \cdot \frac{1}{\delta}, 36(2+c_1)c_1 \cdot \frac{1}{\alpha^2} \cdot \frac{1}{\delta})$. That finishes the proof.

□

3.1.2 Simulations

3.1.2.1 scenario 1

A simulation study was performed to investigate if the formula correct the estimate of Moran's I. The true attribute value x_i for each point p_i on an 20 by 20 grid is generated using R package *gstat*. The weight matrix W is defined so $w_{i,j}$ is $\exp(-\|p_i - p_j\|^2/2(s^2))$, where $\|p_i - p_j\|$ is the Euclidean distance between p_i and p_j and s is a pre-specified scale constant. The observed attribute value y_i is generated by adding a noise ε_i to x_i, and $\epsilon_i \sim N(0, v)$ where $v = f \cdot var(x)$ and f varies from 0.2 to 1 in the simulation. The true Moran's I is calculated by applying formula 1.1 on x_i, and the observed Moran's I is calculated by applying formula 1.1 on y_i, and corrected Moran's I is calculated by applying formula 3.1 on y_i. The simulation is repeated for 100 times for setting of f and mean observed I and corrected I are calculated and presented in table 3.1.

f	0.2	0.4	0.6	0.8	1
true I	0.630	0.630	0.630	0.630	0.630
mean observed I	0.523	0.447	0.392	0.351	0.314
mean corrected I	0.628	0.629	0.630	0.633	0.633

Table 3.1. Experiments for Gaussian noise

Conclusion: The proposed formula corrected the estimate of Moran's I of Gausian distribution.

3.1.2.2 scenario 2

It is common that there are multiple source of variation. Suppose at site i at time point t, there is a population $Y_{i,t} \sim G(\theta_{i,t}, \Theta_i)$ where $\theta_{i,t}$ is the population mean at site i at time point t and Θ_i is the other parameters of the distribution. We also suppose $\theta_{i,t}$ itself is also a random variable and that $\theta_{i,t} \sim N(\theta_i, \sigma_i)$ where θ_i are long term mean of $\theta_{i,t}$. Now we consider the situation that $G(\theta_{i,t}, \Theta_i)$ is Gaussian distribution, that is $Y_{i,t} \sim N(\theta_{i,t}, \Theta_i)$, and we are interested in the spatial correlation of θ_i among all the sites. We need to estimate

θ_i. There are two source of variation of the estimation of θ_i. One is the sampling error, say we can sample a sample of size n_i at time point t to estimate $\theta_{i,t}$, the other is the variation along the time, say we can only do sampling at m time point. However, very often we can make n_i so big that the sampling error is ignorable. Then what we have are $\theta_{i,t}$ which follows Gaussian distribution and the correction mentioned above can also get unbiased estimation of true Moran's I. A simulation is performed to demonstrate this. The true attribute value x_i for each point p_i on an 20 by 20 grid is generated using R package *gstat*. A noise ε_i with known variance v_1 is added to the true value. A sample of size n was drew from Gaussian distribution $N(x_i + \varepsilon_i, v_2)$ from each site and mean m_i of the samples were calculated. Observed and corrected Moran's I were calculated using these m_i. This process was performed for different values of n. The true Moran's I was also calculated using x_i. The results are shown on Figure 3.1.

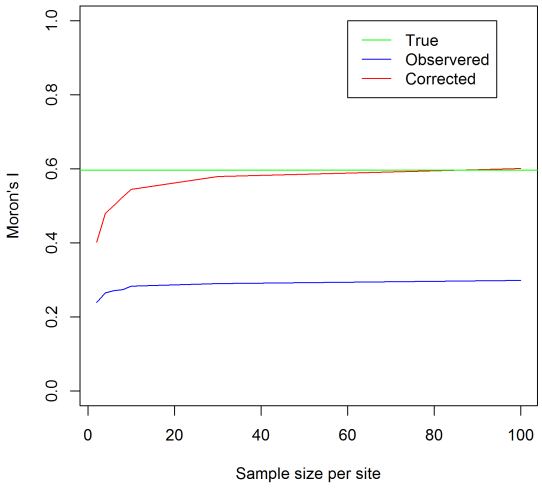

Figure 3.1. The effect of sample size on observed and corrected I in Gaussian distribution

Conclusion: The corrected Moran's I estimate approaches the true value as the sample size increases while the original version of Moran's I is always biased even the sample size is large .

3.2 Application in Non-Gaussian Distribution

So far we assume that the observed attribute value y_i follows Gaussian distribution. Now we investigate if the correction we proposed can also be applied to Non-Gaussian distribution, say Bernoulli distribution, by simulation. In this case the attribute values we are interested in is the probability of event q_i. In the simulation, q_i are restricted in the range $[0.1, 0.5]$. Trial number N_i in each location varies in the simulation. Event count n_i is generated follow binomial distribution with proportion q_i and size N_i and observed proportion q_i^* is then calculated by dividing n_i by N_i. The variance of q_i^* is calculated as $v_i = q_i^* \cdot (1 - q_i^*)/N_i$. Using q_i, q_i^* and v_i as input, true Moran's I, observed Moran's I and corrected Moran's I were calculated similar as above and presented on Figure 3.1.

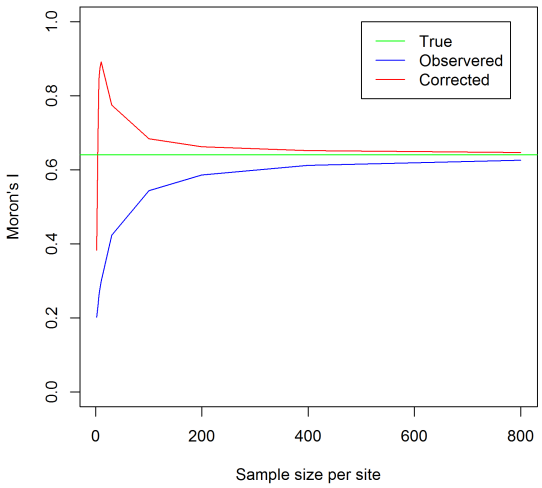

Figure 3.2. The effect of sample size on observed and corrected I in Bernoulli distrubution

Conclusion: The proposed formula also correct the estimate of Moran's I of Bernoulli distribution.

CHAPTER 4

CONCLUSION

During the last decades, the speed of accumulating information has progressed beyond people's imagination. Information is not knowledge however. The first step to gain knowledge from information is to recognize patterns of the information. One of the important patterns is spatial correlation. Moran's I is a commonly used statistic to assess spatial correlation. We provided a sampling method to estimate this statistic and proved bounds of the estimate. We also proposed a corrected formula to estimate Moran's I when there are noise on the attribute values. We proved closeness of the estimate to the true value and demonstrated the closeness property by simulation in 2D space. The future direction can be exploring the performance on the corrected formula on higher dimensional space.

CHAPTER 5

REFERENCE

[1] P. A. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, no. 1/2, pp. 17–23, 1950.

[2] H. Li, C. A. Calder, and N. Cressie, "Beyond moran's i: Testing for spatial dependence based on the spatial autoregressive model," *Geographical Analysis*, vol. 39, no. 4, pp. 357–375, 2007.

[3] M. Helbich, M. Leitner, and N. D. Kapusta, "Geospatial examination of lithium in drinking water and suicide mortality," *International journal of health geographics*, vol. 11, no. 1, p. 19, 2012.

[4] J. Grieve, "A regional analysis of contraction rate in written standard american english," *International Journal of Corpus Linguistics*, vol. 16, no. 4, pp. 514–546, 2011.

[5] M. Maki and M. Masuda, "Spatial autocorrelation of genotypes in a gynodioecious population of chionographis japonica var. kurohimensis (liliaceae)," *International journal of plant sciences*, vol. 154, no. 4, pp. 467–472, 1993.

[6] F. C. Mencken and C. Barnett, "Murder, nonnegligent manslaughter, and spatial autocorrelation in mid-south counties," *Journal of Quantitative Criminology*, vol. 15, no. 4, pp. 407–422, 1999.