

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 15 日現在

機関番号：32601

研究種目：挑戦的萌芽研究

研究期間：2012～2014

課題番号：24650606

研究課題名(和文)「地理の第1法則」を正しく検定する革新的な空間自己相関法の開発

研究課題名(英文) Development of a new spatial auto-correlation method for precisely testing the "First Law of Geography"

研究代表者

岡部 篤行 (Okabe, Atsuyuki)

青山学院大学・総合文化政策学部・教授

研究者番号：10114050

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：本研究は、空間自己相関を検定する統計量として頻繁に使用されているモランのI指標の正確な使用方法を提案したものである。多くの既存研究では、研究対象地域を構成するゾーン数がせいぜい100程度であれば、モランのI指標に正規性が仮定できるとして、空間自己相関がないという帰無仮説の統計的検定が行われている。本研究は、そのような多くのゾーン数であっても正規性は仮定できないことを示し、通常の利用方法は間違いを起こしやすいことを指摘した。この欠点を克服すべく、本研究では膨大な数のモンテカルロシミュレーションを行い、帰無仮説の元での限界値の数表を作成した。この数表を使うことで正確な統計的検定が可能となった。

研究成果の概要(英文)：This study proposes a precise method for using Moran's I index, one of the most frequently used methods for testing spatial autocorrelation. In the related literature, most studies assume that Moran's I index is normally distributed if the number of zones is around one hundred. This study shows that even for that large number, the normality is not satisfied; consequently, empirical studies using Moran's I index with such a number of zones are likely to lead false conclusions. To avoid these false conclusions, this study carried out a huge number of Monte Carlo simulations, and obtained the numerical table of the critical values for Moran's I index. Using this table, researchers can perform a precise statistical test for spatial autocorrelation.

研究分野：地理学

キーワード：空間自己相関 モランのI指標 統計検定用数表 モンテカルロシミュレーション 地理の第1法則

1. 研究開始当初の背景

1970年、地理学で著名な研究者 Tobler は、「地理の第1法則」、すなわち、ものごとは遠いものより近いものからより深く影響を受ける、という経験法則を主張した。以来、この経験法則は、地理学の重要研究課題となり、多くの地理学者が研究を進めてきた。中でも計量地理学者は、この経験法則を空間自己相関法と呼ばれている手法、大まかに言えば空間単位(地区やメッシュ)の隣接する値の相関関係进行分析する手法で実証分析を試みてきた。空間自己相関法は、地理分析教科書では必ず取り上げられ、また多くの専門書も出版されて来ており、さらには S+SpaceStat や Deoda などの GIS ソフトウェアも開発されるなど、地理分析で高い注目を浴びて来た手法である。実際、2004年には米国地理学会学術誌で特集が組まれる程であった。

空間自己相関法で一番使われている指標にモランの I 指標がある(定義は第3節)。この指標が多用されてきた主な理由は、明確な仮定に立脚し、それから演繹される結果は理論的に正しいと確認されており、かつ前提となる仮定は一見、実際にも成り立つように見えたからである。それ故に、実用上も、妥当な方法であると長く信じられて来た。

しかし、筆者らは、本研究に先立つ予備研究で、一見成り立ちそうな仮定の一つが、実は現実的にはほとんど当てはまらないのではないかという疑問を米国地理学会で提起した(学会発表 Yamada and Okabe, 2012 参照)。もしこれが正しいとなると、今まで行われてきたデータ分析の多くの成果は誤っている可能性が高いことになり、これを解明しようということが本研究を始めた背景であった。

2. 研究の目的

本研究の目的は、以下の3つである。

(1) 空間自己相関指標であるモランの I 指標を使い、正規変数空間ランダム法と組み合わせ空間ランダム法(次節参照)が、実際の実証データ分析で使われる各種条件(地区数、どんな確率分布形かなど)に応じて、どの程度に誤った結論を導き出しているかを研究する。

(2) 誤った結論を避ける方法を開発する。

(3) その方法を実装して実用的検定方法を提案する。

3. 研究の方法

研究の方法で取り上げたのは、モランの I 指標である。最初にその定義から述べよう。分析対象地域が N ゾーンから構成されるとし、分析対象の属性値が確率変数 X で与えられ、それぞれのゾーンにおけるその変数の観測値を X_i 、その平均値を \bar{X} 、 w_{ij} をゾーン i と

ゾーン j の近接度合いの値とすると、モランの I は、次の式で与えられる。

$$I = N \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\left(\sum_{i=1}^N \sum_{j=1}^N w_{ij} \right) N \sum_{i=1}^N (X_i - \bar{X})^2}$$

このモランの I 指標のサンプル分布は、空間自己相関がないという帰無仮説の元では、

i) N が十分大きく、

かつ次のどちらかの条件、または二つの条件 ii) 正規分布変数空間ランダム仮定：それぞれのゾーンで観察される $X_i, i=1, \dots, N$ は、同一の正規分布に従って独立に生起する。

iii) 組み合わせ空間ランダム仮定：ゾーンで観察される (X_1, \dots, X_N) は、どの可能な組み合わせも同じ確率で生起する。

を満たすと、漸近的に正規分布に近づくこと証明されている。

もし X に正規分布変数空間ランダム仮定が適用できるとなると、代数的に計算できる平均値と分散値を使って帰無仮説を検定することができる。

もし正規分布変数空間ランダム仮定が適用できないとなると、観測値 (X_1, \dots, X_N) について多くの組み合わせをモンテカルロシミュレーションで発生させてモランの I 指標の近似的なサンプル分布を求め、これを使って帰無仮説を検定することになる。後者は、陰に組み合わせ空間ランダム仮定を置くことを意味するが、実際の検定において、陽に議論されることは少なかった。

本研究においては、実際の分析に使われるデータが正規分布変数空間ランダム仮定を満たすとは限らないことから、組み合わせ空間ランダム仮定での検定方法に焦点を当て、その新たな方法の開発を目指した。すなわち、検定に使われるモランの I 指標の裾部分の分布形を、かつて試みられたことがない程の多くの回数のモンテカルロシミュレーションで観測し、ゾーン数 N と検定精度の実用的な関係を求めた。

ゾーン数 N に着目するのは、仮定 i) の「十分大きく」というのを定量的に捉えたいためである。実証分析研究の文献を見てみると、その値はせいぜい 100 程度であり、多くの文献では、この数が「十分大きい」と想定しているように見えるが、これが果たして正かどうか疑問を持ったからである。

シミュレーション手続きは、以下の通りである。

ステップ1：確率変数 X を設定した所与の同じ確率分布形 D から独立に N 個の値

X_1, \dots, X_N を発生させる。

ステップ2：ステップ1で発生した X_1, \dots, X_N に対してモランの I_{obs} を計算する。

ステップ3： X_1, \dots, X_N の組み合わせをランダ

ムに発生させる。
 ステップ4：ステップ3の組み合わせについてモランのI指標の値 I_k を計算する。
 ステップ5：ステップ3と4をK回繰り返す。その結果得られる $I_k, k=1, \dots, K$ からモランのI指標の分布形を求める。
 ステップ6：ステップ2で得られた I_{obs} をステップ5で得られた分布形を使って上記の帰無仮説の有意性を調べる。
 ステップ7：ステップ1から6までをL回繰り返す。

ステップ2から6は、組み合わせ空間ランダム仮定の検定手続きに対応する。シミュレーションに固有の不安定性を少なくするため、上記の手続きを多く繰り返した。具体的には、 $K=100,000$ 、 $L=1,000$ とした。すなわち、千回の帰無仮説検定を、毎回10万回の組み合わせから得たモランのI指標の分布形を使って行ったということである。

所与の確率分布形Dについては、平均0、分散1の正規分布、平均0、分散1の対数正規分布、平均0、分散0.5の対数正規分布とした。これらの分布形の平均値、分散、歪度、尖度は、表1に示す。対数正規分布を用いたのは、実際のデータには、裾が長く、非対称性の分布が多いからである。

表1：シミュレーションで取り上げた確率分布の理論的モーメント

	Normal distribution $N(0, 1)$	Log-normal distribution $LN(0, 1)$	Log-normal distribution $LN(0, 0.5)$
Mean	0.00	1.65	1.13
Variance	1.00	4.67	0.36
Skewness	0.00	6.18	1.75
Kurtosis	3.00	113.94	8.90

分析対象地域は縦横 $m \times m$ 個の正方形からなる正方形地域とし、 $m=5, 7, 10, 20, 30, 40$ とした。近接度合いの値は、クイーン接続、すなわち、ゾーンiとゾーンjが境界線、頂点を共有する場合に $w_{ij}=1$ 、そうでない場合に $w_{ij}=0$ を用いた。さらに w_{ij} からなるマトリックスに関し、行で標準化 w_{ij}^* を行った。すなわち、

$$w_{ij}^* = w_{ij} / \sum_{j=1}^N w_{ij}$$

と標準化した。

4. 研究成果

最初に、 $K=L=999$ の予備的分析において、モランのI指標が正規分布に従うかどうかを有意水準5%のシャピロ・ウィルク検定で調べてみた。その結果が図1である。

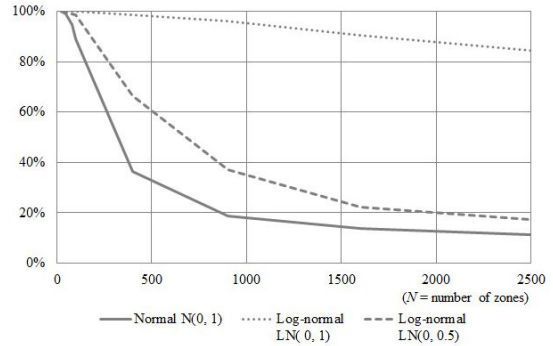


図1：シャピロ・ウィルクテストによるゾーン数と正規分布の割合の関係

この図から分かるようにゾーン数が $50 \times 50 = 2,500$ でも正規分布に従わない割合が10%程度もあることがわかる。

次に5%下限界点についての挙動を図1に示す(なお、5%上限界点、1%上下限界点についても同様な傾向を示している)。限界点の変動の標準偏差は、歪度、尖度が大きくなるにつれ大きくなり、またゾーン数Nが少なくなるほど、大きくなる。変動が無視できるほど小さくなるのはゾーン数Nが大体 30×30 以上の時であり、この数は相当大きく、実際のデータ分析では、これ以下のゾーン数のものが多いので、そのような場合のモランのI指標を利用して分析結果は、再検討を要することが分かる。

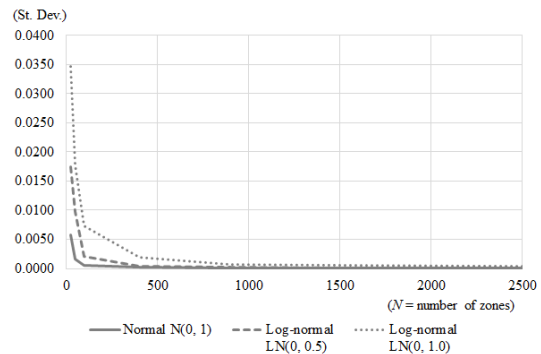


図2：ゾーン数と5%限界点の標準偏差との関係

5%限界点の統計的性質である最大値、最小値、平均と分散は表2のようになる。Theoreticalと記された行は、モランのI指標が正規分布に従うという仮定から導出された5%限界点の値である。この表から読み取れることは、平均値に関し、理論値とシミュレーションの結果の値は、ゾーン数Nがモスほど小さくなって行くが、たとえNが 50×50 の時においても、多くの場合において、差は標準偏差の3倍以上となることが分かる。

表2：シミュレーションにより5%限界点の統計的性質

(a) 平均0、分散1の正規分布の場合

		5x5	7x7	10x10	20x20	30x30	40x40	50x50
Upper 5%	Max	0.2024	0.1483	0.1039	0.0516	0.0339	0.0253	0.0202
	Mean	0.1927	0.1439	0.1018	0.0504	0.0334	0.0249	0.0199
	S.D.	0.0057	0.0016	0.0006	0.0002	0.0002	0.0001	0.0001
	Theoretical	0.1668	0.1299	0.0947	0.0486	0.0326	0.0245	0.0196
Lower 5%	Mean	-0.2227	-0.1570	-0.1076	-0.0518	-0.0340	-0.0253	-0.0201
	Min	-0.2271	-0.1595	-0.1092	-0.0525	-0.0343	-0.0256	-0.0204
	S.D.	0.0026	0.0009	0.0004	0.0002	0.0001	0.0001	0.0001
	Theoretical	-0.2501	-0.1716	-0.1149	-0.0537	-0.0348	-0.0257	-0.0204
Type I errors	Upper (*)	29	25	23	18	30	28	21
	Lower (**)	21	25	23	21	22	27	19
	Total	50	50	46	39	52	55	40

(b) 平均0、分散1の対数正規分布の場合

		5x5	7x7	10x10	20x20	30x30	40x40	50x50
Upper 5%	Max	0.2028	0.1507	0.1193	0.0590	0.0385	0.0282	0.0221
	Mean	0.1675	0.1342	0.1001	0.0520	0.0349	0.0260	0.0207
	S.D.	0.0347	0.0177	0.0074	0.0020	0.0008	0.0005	0.0003
	Theoretical	0.1384	0.1135	0.0866	0.0465	0.0318	0.0240	0.0193
Lower 5%	Mean	-0.1908	-0.1334	-0.0928	-0.0461	-0.0309	-0.0233	-0.0188
	Min	-0.2270	-0.1581	-0.1057	-0.0518	-0.0336	-0.0248	-0.0197
	S.D.	0.0278	0.0180	0.0100	0.0041	0.0018	0.0012	0.0008
	Theoretical	-0.2218	-0.1552	-0.1068	-0.0516	-0.0340	-0.0253	-0.0201
Type I errors	Upper (*)	33	21	28	30	26	16	25
	Lower (**)	32	23	26	25	29	26	22
	Total	65	44	54	55	55	42	47

(c) 平均0、分散0.5の対数正規分布の場合

		5x5	7x7	10x10	20x20	30x30	40x40	50x50
Upper 5%	Max	0.2035	0.1518	0.1043	0.0518	0.0343	0.0255	0.0203
	Mean	0.1856	0.1405	0.1017	0.0509	0.0337	0.0251	0.0200
	S.D.	0.0174	0.0098	0.0021	0.0003	0.0002	0.0001	0.0001
	Theoretical	0.1584	0.1239	0.0927	0.0483	0.0325	0.0244	0.0196
Lower 5%	Mean	-0.2134	-0.1485	-0.1039	-0.0507	-0.0335	-0.0250	-0.0199
	Min	-0.2271	-0.1580	-0.1082	-0.0519	-0.0341	-0.0255	-0.0202
	S.D.	0.0133	0.0109	0.0033	0.0007	0.0002	0.0001	0.0001
	Theoretical	-0.2417	-0.1656	-0.1129	-0.0533	-0.0347	-0.0257	-0.0204
Type I errors	Upper (*)	26	19	19	28	23	28	34
	Lower (**)	30	25	29	17	23	25	18
	Total	56	44	48	45	46	53	52

以上の分析結果をもとに、本研究の目標とその成果の達成度合いをまとめよう。

(1) 研究目的1とその成果

既存文献の実証研究では、空間自己相関を調べるのにモランのI指標が多く利用されている。それを使った統計的検定では、ゾーン数が多い場合、モランのI指標が正規分布に従うということを陽に陰に想定している。しかし、本研究のモンテカルロシミュレーションの結果から、たとえゾーン数が大きい場合でも、モランのI指標が正規分布となることは少ないことが判明した。特に、統計的検定に使用される裾部分において、正規分布を仮定した限界点とモンテカルロシミュレーションから得られる限界点には、ゾーン数がおおくても無視できない差があることが分かり、研究目的1は達成された。

(2) 研究目的2とその成果

所与のゾーン数に対する、モランのI指標の正確な分布形は、変数が正規分布をする場合、理論的に求められるという研究が既に発表されている。しかし、その理論式を実際に計算しようとなると、数値積分で近似をしなければ実用的に求めることができない。また変数が正規分布しない場合の理論式は未だ求

められていない。従って、実用的代替方法は、コンピュータ性能の進歩を利用し、極めて多い回数(100,000)のモンテカルロシミュレーションで検定に使う限界点を求める方法が良いとの結論に達した(研究目的2を達成)。

(3) 研究目的3とその成果

上記の結論から、実用的にはモンテカルロシミュレーションで検定に使う限界点をパラメトリックに求め、それから検定用の数表を策定し、その数表を使って検定するのが正確かつ実用的な方法であるという提案をおこなった(研究目的3を達成)。その一例が表2(a)(b)(c)であり、詳細な数表は公表される。

以上から分かるように、本研究の当初の目的はすべて達成することができたといえよう。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

(雑誌論文)(計2件)

1. Yamada, I., A. Okabe, Examining Tail Distributions of Moran's *I* Statistic through Intensive Simulations. Proceedings of Spatial Data Handling, 2014, 査読有り in press.
2. 山田育穂 岡部篤行 Moranの統計量の裾分布に関する一考察 地理情報学会講演論文集 査読無し 22巻 (CD-Rom) 2013

(学会発表)(計8件)

1. Yamada, I., A. Okabe, Examining Tail Distributions of Moran's *I* Statistic through Intensive Simulations. ISPRS/IGU Joint International Conference on Geospatial Theory, Processing, Modelling and Applications 2014, Toronto, Canada, October 7, 2014.
2. 山田育穂 岡部篤行 解析対象変数の分布形がMoranのI統計量の裾分布に及ぼす影響 日本地理学会秋季学術大会 富山大学, 2014年9月20日
3. Yamada, I., A. Okabe, New Approach to Moran's *I* Test of Spatial Autocorrelation. IGU Kyoto Conference, August 8, 2013.
4. 山田育穂 岡部篤行 Moran I統計量の裾野が空間自己相関の検定に与える影響についての考察 日本地理学会春季大会 国土館大学 2013年3月27日
5. 山田育穂 岡部篤行 MoranのIの統計量を用いた空間自己相関の検定に関する提案 日本地理学会春季大会 立正大学 2013年3月30日
6. Okabe, A., W. Morioka A GIS-based Method for Converting Area-based

Data to Network-based Data and its Application: Estimating the Number of Refugees Who Would Walk to the Nearest Shelter Following a Large Earthquake International Cartographic Conference Dresden August 23, 2013

7. Yamada, L., A. Okabe Toward More Robust Testing of Spatial Autocorrelation by Moran's *I* International Geographic Union IGU Kyoto Regional Conference August 4, 2013.
8. Yamada, L., A. Okabe A note on Moran's *I* statistics for testing spatial autocorrelation The AAG Annual Meeting 2012, New York, NY February 12, 2012

〔図書〕(計0件)

〔産業財産権〕

出願状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

取得状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
取得年月日：
国内外の別：

〔その他〕

ホームページ等

岡部篤行

<http://www2.sccs.aoyama.ac.jp/~atsu/>

山田育穂

<http://www.spatial-lab.human.chuo-u.ac.jp/>

6. 研究組織

(1) 研究代表者

岡部篤行 (OKABE, Atsuyuki)

青山学院大学総合文化政策学部総合文化政策学科・教授

研究者番号：10114050

(2) 研究分担者

山田育穂 (YAMADA, Ikuho)

中央大学理工学部人間総合理工学科・教授

研究者番号：00594756

(3) 連携研究者
なし()
研究者番号：