TESTS of Randomness for Spatial Patterns

Atsushi TAKEUCHI
Matsushita Communication Industrial Co. Ltd.
Saedo-cho Midori-ku, Yokohama, Japan

Hiroe TSUBAKI and Chihiro HIROTSU
Department of Mathematical Engineering,
Faculty of Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

SUMMARY

Two methods of testing for the null hypothesis of randomness are examined. They are the quadrat method and the distance method each based on the efficient score calculated from the negative binomial distribution which is considered as a representative distribution of the clustered pattern.

The quadrat method is essentially the same with the well known variance-to-mean ratio test but its asymptotic null distribution appears to be more accurate when the number of quadrats is moderately large.

The distance method here reduces to the test for exponentiality and is shown to be more powerful than the omnibus Shapiro-Wilk's test.

1. Introduction

Spatial point patterns can be classified largely in three patterns: random pattern, clustered pattern and regular pattern [Diggle (1983)]. A pattern is called random if it is generated from a Poisson process. It is regarded as the most basic pattern since the expected number of points on any unit area in the region is constant and the occurrence of points in any two

Key words: Negative Binomial Model; Quadrat Method; Distance Method; Efficient Score Test.

disjoint area is independent.

In this paper, we investigate the characteristics of two methods, the quadrat method and the distance method, of testing for randomness against the clustered pattern alternative. We assume the negative binomial model as the representative of the clustered pattern [Bliss and Fisher(1953)]. It is well known that the mixture of Poisson, Gamma generates the negative binomial distribution.

When the given pattern is "random", the number of an arbitrary region of unit area follows the Poisson distribution;

$$\Pr (X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (1)$$

with $E[X]=\text{Var}[X]$.

On the other hand, if the number x follows the negative binomial distribution ,

$$\Pr (X = x) = \frac{k (k + 1) \cdots (k + x - 1)}{x!} p^k (1 - p)^x \quad (2)$$

($k > 0$; $0 < p < 1$)

then we have

$$E[x]=k(1-p)/p \quad (3)$$

$$\text{Var}[x]=k(1-p)/p^2 \quad (4)$$

Thus the negative binomial distribution satisfies an important characteristic of the clustered pattern that

$$E[x] < \text{Var}[x].$$

The probability generating function of negative binomial is

$$G(t) = p^k \{ 1 - (1 - p) t \}^{-k} \quad (5)$$

When k tends to infinitive with $E[x]=k(1-p)/p$ fixed at λ , the expression (5) tends to;

$$\begin{aligned} G(t) &= \left(\frac{1}{p} - \frac{1-p}{p} t \right)^{-k} \\ &= \left(\frac{k+\lambda}{k} - \frac{\lambda}{k} t \right)^{-k} \\ &= \left\{ 1 + \frac{\lambda(t-1)}{k} \right\}^{-k} \\ &\rightarrow e^{\lambda(t-1)}, \quad (k \rightarrow \infty) \end{aligned} \quad (6)$$

which is the p.g.f. of the Poisson distribution with parameter λ . So the negative binomial distribution converges to the Poisson distribution with the same mean.

Therefore we can use the parameter " k " as a measure of how the given negative binomial distribution is close to the Poisson distribution.

2. Methods of Testing for Randomness

2-1 Quadrat Methods

One way of testing for randomness of a mapped point pattern is to divide the given region to some subareas and count the number of the points in each subarea. Suppose that the observed counts are N_i for the i -th area ($i=1,2,\dots,m$).

Then the traditional way of testing is based on the sample variance-to-mean ratio statistics [Fisher, Thornton and Mackenzie(1922)].

$$\sum_{i=1}^m \frac{(N_i - N/m)^2}{N/m} \quad (7)$$

which is asymptotically distributed as chi-squared with $m-1$ degrees of freedom under the null hypothesis of randomness.

If we assume the negative binomial model as an alternative, a natural test statistic should be based on the efficient score standardized by the square root of Fisher information, the asymptotic distribution of which is a standardized normal.

$$E = \frac{\sum N_i^2 - N^2 / m + N}{N \sqrt{m / 2}} \sim N(0, 1) \quad (8)$$

These two statistics are essentially the same, excepting a possible difference in the accuracy of the chi-square approximation.

2-2 Distance Methods

Another way of testing is to examine the distances of "tree-tree" ("tree" is a data point) or "point-tree" ("point is a random point in the area") [Ripley(1982)]. Among them, the nearest neighborhood distance is usually used. Since this distance follows exponential distribution under the null hypothesis of randomness, the required test reduce to a test for exponentiality. For this purpose we can use the "Shapiro-Wilk's W";

$$W = \frac{m (\bar{x} - x_{min})^2}{(m-1) \sum (x_i - \bar{x})^2} \quad (9)$$

In (9), "m" is the number of random points, and x_i ($i=1,2,\dots,m$) is the area of the circle which has the radius of the nearest

neighbor distance from the i -th random point. And x is the sum of x_i/m .

Again if the negative binomial model is assumed as an alternative, we can consider a test based on the efficient score;

$$E = \left\{ \frac{\sum (x_i - \bar{x})^2}{2 \bar{x}^2} - m/2 \right\} / \sqrt{m} \quad (10)$$

Expression (10) is derived from the fact that, if the given pattern is a Poisson-Gamma mixture, the distribution of a nearest-neighbor area of a circle generates the Pareto distribution [Eberhardt(1967)]. The essentially equivalent statistics was proposed by Brown and Rothery(1978), who did not suggest explicitly that the test based on (10) is locally most powerful.

3. Power of Test in These Methods

3-1 Asymptotic Power of the Variance-to-Mean Ratio test and Its Simulation Study

In case of the negative binomial distribution alternative we can evaluate the asymptotic power of the variance-to-mean ratio statistics as follows. Let k, p be the parameters of the negative binomial distribution and N the number of observed points in the given area consisting of m subareas, then the asymptotic power is given by

$$\Pr \left\{ \frac{m}{m-k+1} X \geq \chi_{\alpha}^2 \right\} = \int_{\frac{mk+1}{mk+N}}^{\infty} \frac{1}{2 \Gamma(\frac{m-1}{2})} (x/2)^{(m-3)/2} e^{-x/2} dx \quad (11)$$

$$(X \sim \chi^2(m-1))$$

Figure 1 is the asymptotic power of this method calculated by

numerical integration of (11) (the significance level is $\alpha=0.05$).

Figure 1 also shows the simulation study (1000 times) for the power of this test against negative binomial pattern.

Fig.1-a N=100

Fig.1-b N=200

Fig. 1 Asymptotic Power of the Quadrat Method

3-2 Chi-square Approximation And Normal Approximation

The variance-to-mean ratio statistic and the efficient score statistic are essentially the same. Therefore we compare the accuracy of the significant level in order to compare these two methods.

Table 1 shows the observed significance level by the simulation of 1000 times when the official significance level is fixed at 0.05 .

Table 1 Comparison of Observed Significance Levels of the Variance-to-Mean Ratio Test and the Test based on Efficient Score

Table 1-a N=100			Table 1-b N=200		
m	32	64	m	32	64
V/M ratio test	0.041	0.024	V/M ratio test	0.034	0.032
Eff. score test	0.044	0.040	Eff. Score test	0.042	0.050

3-3 Simulation for Distance Methods

In order to execute an accurate simulation for distance methods, we must form a real area so that many data points from the Poisson-Gamma mixture are included for generating data of distances. But this method takes too much time for simulation. So we generate data of areas of circles by Pareto distribution or exponential distribution, and regard them as data derived from a mapped data.

Figure 2 shows the estimated power curve of the Shapiro-Wilk's test and the efficient score test by our simulation.

Further we study the accuracy of the Normal approximation of the efficient score statistics by an observed significance level. Table 2 shows the observed significance level of the efficient score test by the simulation of 1000 times where the official is fixed at 0.05 .

Fig.2-a S-W's W

Fig.2-b efficient

Fig. 2 Power Comparison between S-W Test and Eff. Score Test

Table 2 Accuracy of the Normal Approximation
for the Null Distribution of the Efficient Score Test

m	16	32	64	128	256
Obs. Significance Level	0.015	0.027	0.034	0.038	0.043

4. Discussions

4-1. Quadrant Methods

Efficient score test is equivalent to the variance-to-mean ratio test in principle. The simple normal approximation of the efficient score test is, however, recommended since the variance-to-mean ratio test appears often too conservative for moderate number of quadrats.

4-2. Distance Methods

The efficient score test in the distance method is shown as powerful as the Shapiro-Wilk's test by our simulation. and our test is much more conservative than the Shapiro-Wilk's test so that the power of our test will be fairly increased if the exact null distribution of the test statistics can be used, which is partially verified by Brown and Rothery(1978) with Monte-Carlo simulation. The result is not surprising since the Shapiro-Wilk's test is an omnibus test for exponentiality.

ACKNOWLEDGMENT

The authors would like to thank Dr. Y. Ohashi for several suggestions which have substantially improved our study.

REFERENCES

- Bliss, C. I. and Fisher, R. A. [1953]. Fitting the negative binomial distribution to biological data and note on the efficient fitting of the negative binomial. Biometrics 9, 176-200.
- Brown, D. and Rothery P. [1978]. Randomness and local regular points in a plane. Biometrika 65, 115-122.
- Diggle, P. J. [1983]. Statistical analysis of spatial point patterns. Academic Press.
- Eberhardt, L. L. [1967]. Some developments in 'distance sampling'. Biometrics 23, 207-216.
- Fisher, R. A., Thornton, H. G. and Mackenzie, W. A. [1922]. The accuracy of the plating method of estimating the density of bacterial populations, with particular reference to the use of Thornton's agar medium with soil samples. Ann. appl. Bot. 9, 325-359.
- Ripley, B. D. [1981]. Spatial Statistics. John Wiley & Sons.