12-31-2006

Correction to Koen's critical values in testing spatial randomness

Sung Nok Chiu

Hong Kong Baptist University, snchiu@hkbu.edu.hk

Correction to Koen's critical values in testing spatial randomness

Sung Nok Chiu

Department of Mathematics, Hong Kong Baptist University,
Kowloon Tong, Hong Kong
snchiu@hkbu.edu.hk

Abstract

Using Ripley's K-function in testing spatial randomness, Koen's estimated critical values deviated notably from Ripley's approximation formula; the former, however, came from an incorrect algorithm. This paper reports new estimates, which agree very well with Ripley's approximation and recommends that the approximation formula can be used instead of Monte-Carlo tests.

Keywords: complete spatial randomness, spatial point patterns.

1 Introduction

Given a spatial point pattern on $\mathbb{R}^2$, we typically start an analysis with a test of the complete spatial randomness (CSR) hypothesis, because 'rejection of CSR is a minimal prerequisite to any serious attempt to model an observed pattern' (Diggle [3], p. 12). One commonly used statistic for testing the CSR hypothesis is the maximum absolute pointwise difference between the theoretical form of a summary statistic under the CSR hypothesis and its estimator. The distribution of such a statistic is typically unknown but of course can be approximated by Monte-Carlo simulation.

In particular, if we consider Ripley's [9] K-function, defined by

$$K(r) = \frac{\text{Average number of further points within distance } r \text{ of an arbitrary point}}{\lambda}, \quad r \geq 0,$$

where λ is the intensity of the point process, the theoretical form of it under the CSR hypothesis is simply $K(r) = \pi r^2$.

Given a point pattern observed in a bounded sampling window W, the estimation of K is hampered by the edge-effects. Suppose there are n points observed in W of area A. Let u_{ij} be the distance between points i and j. The empirical mean $\sum_{i=1}^{n} \sum_{i \neq j}^{n} \mathbf{1}_{[0,r]}(u_{ij})/n$ underestimates the true $\lambda K(r)$, because some unobserved points lying close to but outside W may be within distance r of an observed point lying close to the boundary of W. One commonly used edge-correction is Ripley's [9] isotropic correction:

$$\hat{K}_{\text{iso}}(r) = \frac{\sum_{i=1}^{n}\sum_{i\neq j}^{n} w_{ij}^{-1}\mathbf{1}_{[0,r]}(u_{ij})}{n^2},$$

where w_{ij} is the proportion of circumference of the circle centred at point i with radius u_{ij} lying in W. Some authors (e.g. Diggle [3], p. 50) use $n(n-1)$ rather than n^2 in the denominator for technical reasons. Even better is to use the adapted distance dependent intensity estimators proposed by Stoyan and Stoyan [13]. Exact and asymptotic formulae for the variance of $\hat{K}_{\text{iso}}(r)$ have been derived by Chetwynd and Diggle [1], Lotwick and Silverman [7] and Ripley [11] (pp. 39-40). As for testing CSR, we would like to know the distribution of

$$\tau = \sup_{r\leq r_0}\left|\sqrt{\frac{\hat{K}_{\text{iso}}(r)}{\pi}} - r\right|.$$

Koen [6] used simulation to estimate the 90th, 95th and 99th percentile of τ, as a function of the number of points n and the arbitrary upper limit r_0, for the special case that W is a unit square. These values were reproduced in Stoyan and Stoyan [12] and still were used in, e.g., Mateu [8] after ten years of the publication. Koen's estimates of the 95th percentile deviate substantially from Ripley's [11] approximation formula, which says that if nUr_0^3/A^2 is small, where U is the perimeter of W, then the 95th percentile is approximately $1.45\sqrt{A}/n$ for a wide range of r_0. In this short note we report the error in Koen's algorithm and show that Ripley's formula in fact gives very good approximates for the 95th percentile.

2 Simulation

In the calculation of $\hat{K}_{\text{iso}}(r)$, we have to construct a circle centred at each point with radius r. Koen's algorithm considers two scenarios: such a circle intersects (i) only one side of the square W and (ii) two sides of W. In the second scenario, however, he only had a formula for the situation that the distance, say d, between the centre and the corner formed by the two sides of W intersecting the circle is not greater than r, so that the corner is contained in the closed circle. For $r \leq 0.5$, in fact, we have to distinguish two cases: (i) $r \leq d$, and (ii) $r > d$ (see Diggle [3], p. 51), the latter has obviously been missed by Koen. For $r > 0.5$, more scenarios have to be considered (Haase [4]). Nevertheless, various libraries in `R`, such as `spatial`, `spatstat` and `splances`, already incorporated more advanced algorithms for computing w_{ij} for polygonal W.

Table 1 shows the estimates of the 90th, 95th and 99th percentiles of τ, based on 20,000 simulated realisations of a binomial process in a unit square, generated by the default random number generator (Mersenne–Twister) in `R` version 2.2.1. They differ remarkably from Koen's estimates. The products of the percentiles and n were shown in Table 2, from which we can see that when $r_0 = 1.25/\sqrt{n}$, the value recommended by Ripley [10] for the upper limit, the agreement between Ripley's approximation formula $1.45\sqrt{A}/n$ and the estimated 95th percentile is very good. For other values of r_0, however, the factor 1.45 has to be changed. Nevertheless, the power comparison in Ho and Chiu [5] shows empirically that using $r_0 = 1.25/\sqrt{n}$ in τ, though not necessarily, often leads to a more powerful test than using larger r_0. Thus, if we adopt this recommended upper limit, then we may simply use $1.31\sqrt{A}/n$, $1.45\sqrt{A}/n$ and $1.75\sqrt{A}/n$ to approximate the critical values at the 0.1, 0.05 and 0.01 significance levels, respectively, instead of using Monte-Carlo tests (Diggle [2]). The corresponding estimated type I error rates at different n are tabulated in Table 3.

Table 1 about here

Table 2 about here

Table 3 about here

Acknowledgements

This research was partially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Numbers HKBU2048/02P and HKBU200503) and an FRG grant of the Hong Kong Baptist University.

References

[1] Chetwynd, A. G. and Diggle, P. J., 1998, On estimating the reduced second moment measure of a stationary spatial point process. *Australian and New Zealand Journal of Statistics*, **40**, 11-15.

[2] Diggle, P. J., 1979, On parameter estimation and goodness-of-fit testing for spatial point patterns. *Biometrics*, **35**, 87-101.

[3] Diggle, P. J., 2003, *Statistical Analysis of Spatial Point Patterns* (2nd edn) (London: Arnold).

[4] Haase, P., 1995, Spatial pattern analysis in ecology based on Ripley's K-function: Introduction and methods of edge correction. *Journal of Vegetation Science*, **6**, 575-582.

[5] Ho, L. P. and Chiu, S. N., 2006, Testing the complete spatial randomness by Diggle's test without an arbitrary upper limit. *Journal of Statistical Computation and Simulation*, **76**, 585-591.

[6] Koen, C., 1991, Approximate confidence bounds for Ripley's statistic for random points in a square. *Biometrical Journal*, **33**, 173-177.

[7] Lotwick, H. W. and Silverman, B. W., 1982, Methods for analysing spatial processes of several type of points. *Journal of the Royal Statistical Society Series B*, **44**, 406-413.

[8] Mateu, J., 2001, Parametric procedures in the analysis of replicated pairwise interaction point patterns. *Biometrical Journal*, **43**, 375-394.

[9] Ripley, B. D., 1976, The second-order analysis of stationary point processes. *Journal of Applied Probability*, **13**, 255-266.

[10] Ripley, B. D., 1979, Tests of 'randomness' for spatial point patterns. *Journal of the Royal Statistical Society Series B*, **41**, 368-374.

[11] Ripley, B. D., 1988, *Statistical Inference for Spatial Processes* (Cambridge: Cambridge University Press).

[12] Stoyan, D. and Stoyan, H., 1994, *Fractals, Random Shapes and Point Fields* (Chichester: John Wiley & Sons).

[13] Stoyan, D. and Stoyan, H., 2000, Improving ratio estimators of second order point process characteristics. *Scandinavian Journal of Statistics*, **27**, 641-656.

Table 1: Estimates of the 90th, 95th and 99th percentiles of τ from 20,000 simulations. Koen's estimates are given in brackets.

		r_0		
percentile	n	$0.75/\sqrt{n}$	$1.25/\sqrt{n}$	$2.5/\sqrt{n}$
90th	10	0.13215 (0.0843)	0.13708 (0.104)	—
	20	0.06491 (0.0559)	0.06646 (0.0559)	—
	25	0.05179 (0.05)	0.05307 (0.05)	0.05717 (0.05)
	30	0.04302 (0.0326)	0.04402 (0.0362)	0.04698 (0.0430)
	40	0.03229 (0.0210)	0.03295 (0.0237)	0.03464 (0.0280)
	50	0.02580 (0.0187)	0.02630 (0.0194)	0.02749 (0.0213)
	100	0.01289 (0.00885)	0.01309 (0.00918)	0.01353 (0.0104)
	200	0.00646 (0.00434)	0.00655 (0.00465)	0.00671 (0.00509)
	300	0.00432 (0.00278)	0.00437 (0.00297)	0.00470 (0.00334)
95th	10	0.14834 (0.113)	0.15201 (0.124)	—
	20	0.07260 (0.0559)	0.07385 (0.0559)	—
	25	0.05780 (0.05)	0.05881 (0.05)	0.06262 (0.0509)
	30	0.04813 (0.0456)	0.04888 (0.0456)	0.05144 (0.0456)
	40	0.03598 (0.0259)	0.03650 (0.0284)	0.03797 (0.0315)
	50	0.02874 (0.0198)	0.02915 (0.0213)	0.03014 (0.0245)
	100	0.01430 (0.00986)	0.01450 (0.0108)	0.01489 (0.0114)
	200	0.00717 (0.00495)	0.00724 (0.00527)	0.00739 (0.00567)
	300	0.00479 (0.00313)	0.00482 (0.00339)	0.00491 (0.00365)
99th	10	0.17941 (0.158)	0.18144 (0.158)	—
	20	0.08801 (0.0605)	0.08884 (0.0665)	—
	25	0.07002 (0.0511)	0.07067 (0.0537)	0.07403 (0.0629)
	30	0.05838 (0.0456)	0.05886 (0.0456)	0.06410 (0.0509)
	40	0.04351 (0.0395)	0.04393 (0.0395)	0.04514 (0.0395)
	50	0.03483 (0.0253)	0.03508 (0.0263)	0.03605 (0.0299)
	100	0.01739 (0.0131)	0.01750 (0.0137)	0.01781 (0.0138)
	200	0.00872 (0.00588)	0.00875 (0.00681)	0.00888 (0.00729)
	300	0.00580 (0.00380)	0.00581 (0.00396)	0.00587 (0.00465)

Table 2: Products of the percentiles and the number of points — Ripley's factor

	90th percentile			95th percentile			99th percentile		
	$r_0\sqrt{n}$			$r_0\sqrt{n}$			$r_0\sqrt{n}$		
n	0.75	1.25	2.5	0.75	1.25	2.5	0.75	1.25	2.5
10	1.3215	1.3708	—	1.4834	1.5201	—	1.7941	1.8144	—
20	1.2982	1.3292	—	1.4521	1.4770	—	1.7601	1.7767	—
25	1.2948	1.3268	1.4292	1.4450	1.4702	1.5655	1.7504	1.7668	1.8507
30	1.2907	1.3207	1.4095	1.4439	1.4663	1.5433	1.7513	1.7658	1.9229
40	1.2914	1.3180	1.3854	1.4390	1.4600	1.5189	1.7403	1.7570	1.8055
50	1.2899	1.3152	1.3743	1.4368	1.4573	1.5080	1.7413	1.7541	1.8026
100	1.2887	1.3091	1.3532	1.4303	1.4498	1.4885	1.7385	1.7500	1.7813
200	1.2919	1.3097	1.3423	1.4348	1.4484	1.4780	1.7448	1.7507	1.7765
300	1.2962	1.3120	1.3410	1.4355	1.4472	1.4726	1.7393	1.7437	1.7602

Table 3: Estimates of the type I error rate at $r_0 = 1.25/\sqrt{n}$

Ripley's	n							
factor	10	20	30	40	50	100	200	300
1.31	0.13073	0.11084	0.10503	0.10393	0.10209	0.10030	0.09988	0.10091
1.45	0.06968	0.05751	0.05429	0.05258	0.05139	0.05024	0.04961	0.04927
1.75	0.01434	0.01166	0.01104	0.01037	0.01013	0.00995	0.01005	0.00969