Journal of the Operations Research Society of Japan
Vol. 55, No. 2, June 2012, pp. 146–157

AN APPROXIMATION FOR THE KTH NEAREST DISTANCE AND ITS APPLICATION TO LOCATIONAL ANALYSIS

Masashi Miyagawa
University of Yamanashi

(Received May 6, 2011; Revised March 12, 2012)

Abstract In this paper, we provide a simple approximation for the distance from an arbitrary location to the kth nearest point. Distance is measured as the Euclidean and the rectilinear distances on a continuous plane. The approximation demonstrates that the kth nearest distance is proportional to the square root of k and inversely proportional to the square root of the density of points. The accuracy of the approximation is assessed for regular and random point patterns. Comparing the approximation with road network distances shows that the approximation on a continuous plane can be used for estimating the kth nearest distance on actual road networks. As an application of the approximation to locational analysis, we obtain the average distance to the nearest open facility when some of the existing facilities are closed.

Keywords: Facility planning, point pattern, Euclidean distance, rectilinear distance, road network distance

1. Introduction

The nearest neighbour distance, which is the distance between neighbouring points, has been used in point pattern analysis for describing patterns for the distribution of various geographical objects [4, 12]. Since the nearest neighbour distance method was introduced by Clark and Evans [3], many statistics based on the nearest neighbour distance have been proposed. Although the nearest neighbour distance is the most important, the distance to the kth nearest point is necessary to deal with complicated patterns. The distance to the second nearest point was considered in Holgate [11]. Ripley's K-function, which is one of the most frequently used tools for point pattern analysis, handles distances between all pairs of points as well as the kth nearest distance [25]. The K-function method has been applied to the distribution of population [8], traffic accidents [13], and trees [9].

Another application of the nearest neighbour distance can be found in locational analysis. The distance from customers to their nearest facility represents the service level of facility location. The distance to the kth nearest facility is also important when facilities are closed or disrupted because of accidents and disasters. A number of facility location models incorporating a reliability aspect have considered non-closest facility service. Weaver and Church [30] addressed the vector assignment p-median problem, where a certain percentage of customers could be serviced by the kth nearest facility. Pirkul [24] studied a similar problem in which customers are served by two facilities designated as primary and secondary facilities. Drezner [6] formulated the unreliable p-median and p-centre problems and suggested heuristic solutions when the probability of facility failure is the same for all facilities. In both models, customers are assigned to the kth nearest facility when closer facilities have failed. Berman et al. [2] extended the Drezner's model by assuming that the probabilities of facility failure are not identical. Snyder and Daskin [26] presented two reliability models based on the p-median problem and the uncapacitated fixed-charge location problem. They made an ordered assignment of each customer to each facility. Non-closest facility service is also found in emergency vehicle location models, where the service availability is computed using queueing theory [16, 18, 27].

Analytical expressions for the kth nearest distance have been obtained for regular and random point patterns. An overview of the literature is presented in Table 1. The nearest Euclidean distance was derived in Clark and Evans [3] for the random pattern, Persson [23] for the square lattice, and Holgate [10] for the triangular lattice. The kth nearest Euclidean distance was derived in Thompson [28] and Dacey [5] for the random pattern, Koshizuka [14] for $k = 1, 2, 3$ for the square lattice, and Miyagawa et al. [21] for $k = 1, 2, \ldots, 7$ for the square, triangular, and hexagonal lattices. The kth nearest rectilinear distance has also been obtained. The nearest rectilinear distance was derived in Larson and Odoni [17] for the random pattern. The kth nearest rectilinear distance was derived in Miyagawa [19] for the random pattern, and for $k = 1, 2, \ldots, 8$ for the square and diamond lattices. For higher order distances of regular patterns, upper and lower bounds were obtained [19–21].

Table 1: Literature on the kth nearest distance

	Year	Distance	k	Pattern
Clark and Evans [3]	1954	Euclidean	1	Random
Persson [23]	1964	Euclidean	1	Square
Holgate [10]	1965	Euclidean	1	Triangular
Thompson [28]	1956	Euclidean	$1, \ldots, \infty$	Random
Dacey [5]	1968	Euclidean	$1, \ldots, \infty$	Random
Koshizuka [14]	1985	Euclidean	$1, 2, 3$	Square
Miyagawa et al. [21]	2004	Euclidean	$1, \ldots, 7$	Square, Triangular, Hexagonal
Larson and Odoni [17]	1981	Rectilinear	1	Random
Miyagawa [19]	2008	Rectilinear	$1, \ldots, \infty$	Random
		Rectilinear	$1, \ldots, 8$	Square, Diamond

In this paper, we provide an approximation for the kth nearest distance. The kth nearest distance is defined as the distance from an arbitrary location to the kth nearest point. The approximation allows us to estimate the kth nearest distance that has not previously been derived. Given the distribution of points, the kth nearest distance can be numerically calculated. The result, however, depends on the specific data and cannot be applied to other situations. The approximation is useful to examine fundamental relationships between variables, for example, how the density of points affects the kth nearest distance. Thus, the approximation will supply building blocks for further spatial analysis based on the kth nearest distance.

The remainder of this paper is organized as follows. The next section obtains an approximation of the Euclidean distance. The accuracy of the approximation is then assessed for regular and random point patterns. Section 3 gives an approximation of the rectilinear distance. Section 4 compares the approximation with the kth nearest distance on a road network. Section 5 provides an application of the approximation to locational analysis. The final section presents concluding remarks.

2. Euclidean Distance

Let R_k be the Euclidean distance from an arbitrary location in a study region to the kth nearest point. We call R_k the kth nearest distance. In this section, we give an approximation for the kth nearest distance R_k.

Consider a circle centred at an arbitrary location with radius R_k. The circle contains k points (one point on the circumference and $k-1$ points in the inside). The number of points in the circle can be approximated as

$$k \approx \rho\pi R_k^2, \tag{2.1}$$

where ρ is the density of points (the number of points per unit area). An approximation for R_k is then obtained as

$$R_k \approx \sqrt{\frac{k}{\rho\pi}}. \tag{2.2}$$

For a deeper understanding of this approximation, we examine R_k of regular and random point patterns.

2.1. Grid pattern

Suppose that points are regularly distributed on a square grid with side length a. The grid pattern is assumed to continue infinitely. This assumption enables us to examine the kth nearest distance without taking into account the boundary effect. Let R_k be the Euclidean distance from an arbitrary location to the kth nearest point distributed regularly. Let us consider the relationship between the area of a circle with radius R_k and the area of squares centred at grid points within the circle, as illustrated in Figure 1. The area of the circle is πR_k^2, whereas the area of the squares is ka^2, because the circle contains k points. The difference in these areas is at most the area of squares that intersect the circumference. Since the squares that intersect the circumference lie between two concentric circles with radii $R_k - \sqrt{2}a$ and $R_k + \sqrt{2}a$, we have

$$\pi(R_k - \sqrt{2}a)^2 < ka^2 < \pi(R_k + \sqrt{2}a)^2. \tag{2.3}$$

Solving for R_k yields

$$\left(\sqrt{\frac{k}{\pi}} - \sqrt{2}\right) a < R_k < \left(\sqrt{\frac{k}{\pi}} + \sqrt{2}\right) a, \tag{2.4}$$

which reduces to

$$1 - \sqrt{\frac{2\pi}{k}} < R_k \Bigg/ \sqrt{\frac{k}{\rho\pi}} < 1 + \sqrt{\frac{2\pi}{k}}, \tag{2.5}$$

where ρ $(= 1/a^2)$ is the density of points. Then we have

$$R_k \Bigg/ \sqrt{\frac{k}{\rho\pi}} \to 1 \qquad (k \to \infty). \tag{2.6}$$

The approximation (2.2) therefore corresponds to the asymptotic value as k tends to infinity.

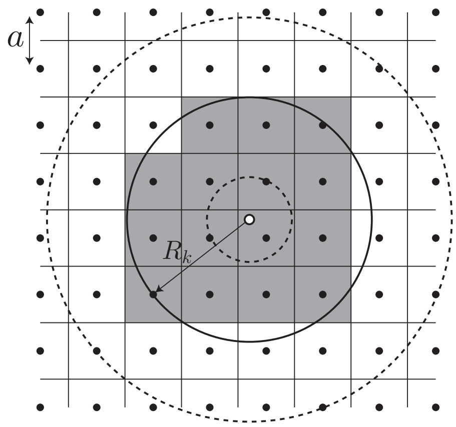


Figure 1: Circle with radius R_k and points within the circle

2.2. Random pattern

Suppose that points are uniformly and randomly distributed on a plane. The random pattern is assumed to continue infinitely. Let R_k be the Euclidean distance from an arbitrary location to the kth nearest point distributed at random. The average kth nearest distance $E(R_k)$ was obtained in Thompson [28] and Dacey [5] as

$$E(R_k) = \frac{(2k-1)!!}{(2k-2)!!}\frac{1}{2\sqrt{\rho}}, \tag{2.7}$$

where $(2k-1)!! = (2k-1)(2k-3)\cdots 5 \cdot 3 \cdot 1$, which reduces to

$$E(R_k) = \frac{(2k-1)!!}{(2k)!!}\frac{k}{\sqrt{\rho}}. \tag{2.8}$$

From Wallis' formula

$$\lim_{k\to\infty} \frac{(2k)!!}{(2k-1)!!}\frac{1}{\sqrt{k}} = \sqrt{\pi}, \tag{2.9}$$

we have

$$E(R_k) \Big/ \sqrt{\frac{k}{\rho\pi}} \to 1 \qquad (k \to \infty). \tag{2.10}$$

Note that the asymptotic value of the random pattern is identical with that of the grid pattern. Hence, the difference in the kth nearest distance between the two patterns is small for large k. The average distance of the random pattern (2.7) might also be used as an approximation for the kth nearest distance. Our approximation (2.2) is, however, simpler than (2.7).

Now, let us assess the accuracy of the approximation. The average distance $E(R_k)$ ($k = 1, 2, \ldots, 7$) of the grid and random patterns and the approximation (2.2) are shown in Figure 2, where the density of points is $\rho = 1$. The relative error of the approximation is shown in Table 2. Although the error is somewhat large for small k, the approximation becomes more accurate as k increases.

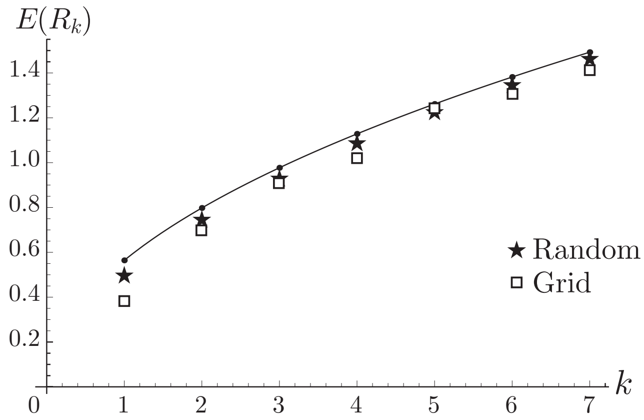


Figure 2: Average kth nearest Euclidean distance

Table 2: Relative error of the approximation (%)

	$E(R_1)$	$E(R_2)$	$E(R_3)$	$E(R_4)$	$E(R_5)$	$E(R_6)$	$E(R_7)$
Grid	47.5	14.1	7.6	10.3	1.5	5.5	5.7
Random	12.8	6.4	4.2	3.2	2.5	2.1	1.8

3. Rectilinear Distance

Although the Euclidean distance is a good approximation for the actual travel distance, the rectilinear distance is more suitable for cities with a grid road network [29]. In fact, the rectilinear distance has frequently been used in facility location models [1, 7, 22]. The rectilinear distance between two points $(x_1, y_1), (x_2, y_2)$ is defined as $|x_1 - x_2| + |y_1 - y_2|$. Let R_k be the rectilinear distance from an arbitrary location in a study region to the kth nearest point. In this section, we give an approximation for the kth nearest rectilinear distance R_k.

An approximation for R_k can be obtained by the same way as the Euclidean distance case, except for replacing a circle with a diamond, which is a square rotated at angle $\pi/4$. Recall that a diamond gives the set of points within a given rectilinear distance from its centre (see, e.g., [15]). Since the area of a diamond with radius R_k is $2R_k^2$, the number of points in the diamond can be approximated as

$$k \approx 2\rho R_k^2. \tag{3.1}$$

An approximation for R_k is then obtained as

$$R_k \approx \sqrt{\frac{k}{2\rho}}. \tag{3.2}$$

The approximation of the rectilinear distance (3.2) is $\sqrt{\pi/2}$ (≈ 1.25) times as large as that of the Euclidean distance (2.2). This result is consistent with the fact that the ratio of the average kth nearest rectilinear distance to the average kth nearest Euclidean distance is $\sqrt{\pi/2}$ for the random pattern.

3.1. Grid pattern

Let R_k be the rectilinear distance from an arbitrary location to the kth nearest point distributed regularly. Let us consider the relationship between the area of a diamond with radius R_k and the area of squares centred at grid points within the diamond, as illustrated in Figure 3. The difference in these areas is at most the area of squares that intersect the circumference. Since the squares that intersect the circumference lie between two concentric diamonds with radii $R_k - 2a$ and $R_k + 2a$, we have

$$2(R_k - 2a)^2 < ka^2 < 2(R_k + 2a)^2. \tag{3.3}$$

Solving for R_k yields

$$\left(\sqrt{\frac{k}{2}} - 2\right) a < R_k < \left(\sqrt{\frac{k}{2}} + 2\right) a, \tag{3.4}$$

which reduces to

$$1 - \frac{2\sqrt{2}}{\sqrt{k}} < R_k \Big/ \sqrt{\frac{k}{2\rho}} < 1 + \frac{2\sqrt{2}}{\sqrt{k}}, \tag{3.5}$$

where ρ $(= 1/a^2)$ is the density of points. Then we have

$$R_k \Big/ \sqrt{\frac{k}{2\rho}} \to 1 \qquad (k \to \infty). \tag{3.6}$$

The approximation (3.2) therefore corresponds to the asymptotic value as k tends to infinity.

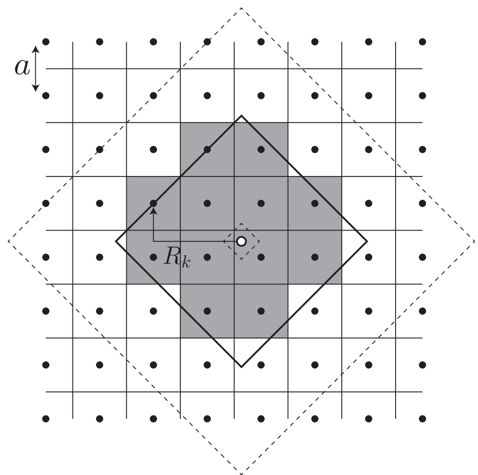


Figure 3: Diamond with radius R_k and points within the diamond

3.2. Random pattern

Let R_k be the rectilinear distance from an arbitrary location to the kth nearest point distributed at random. The average kth nearest distance $E(R_k)$ was obtained in Miyagawa [19] as

$$E(R_k) = \frac{(2k-1)!!}{(2k-2)!!} \frac{\sqrt{\pi}}{2\sqrt{2\rho}}. \tag{3.7}$$

From Wallis' formula (2.9), we have

$$E(R_k) \Big/ \sqrt{\frac{k}{2\rho}} \to 1 \qquad (k \to \infty). \tag{3.8}$$

The average distance $E(R_k)$ ($k = 1, 2, \ldots, 8$) of the grid and random patterns and the approximation (3.2) are shown in Figure 4, where the density of points is $\rho = 1$. The relative error of the approximation is shown in Table 3. The approximation becomes more accurate as k increases.

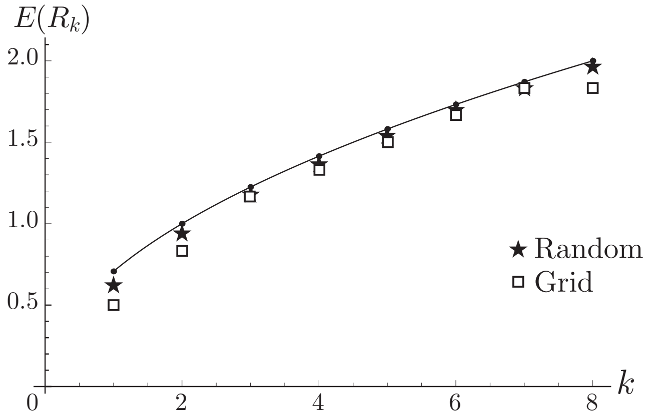


Figure 4: Average kth nearest rectilinear distance

Table 3: Relative error of the approximation (%)

	$E(R_1)$	$E(R_2)$	$E(R_3)$	$E(R_4)$	$E(R_5)$	$E(R_6)$	$E(R_7)$	$E(R_8)$
Grid	41.4	20.0	5.0	6.1	5.4	3.9	2.0	9.1
Random	12.8	6.4	4.2	3.2	2.5	2.1	1.8	1.6

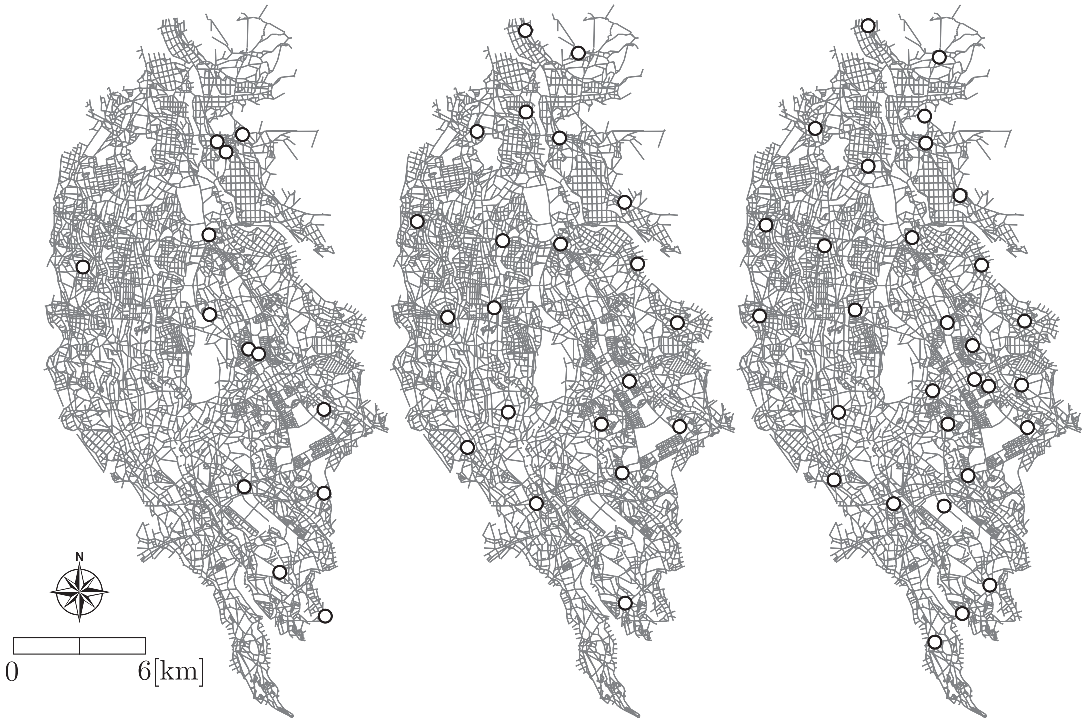
4. Road Network Distance

In this section, we discuss whether or not the approximation on a continuous plane can be used for estimating the kth nearest distance on actual road networks. As an example, we consider road network distances to hospitals, police stations, and post offices in Tsukuba, Japan, as shown in Figure 5.

Let R_k be the road network distance from an arbitrary node of the network to the kth nearest facility. The average kth nearest distance $E(R_k)$ is shown in Figure 6. The solid and dotted curves are the approximations of the Euclidean distance (2.2) and the rectilinear distance (3.2), respectively. The difference between the road network distance and the approximation increases with k. This is because we consider a finite network, even though the approximation is based on an infinite plane. Apart from this boundary effect, there exists a close relationship. Obviously, this may not be the case for clustering patterns of facilities. If facilities are clustered, the distance to the nearest facility would be much greater than the approximation. As long as facilities are sufficiently dispersed, however, the approximation on a continuous plane gives a good estimate for the road network distance to the kth nearest facility.

5. Application

The kth nearest distance is useful for facility location problems with closing of facilities, as discussed in Miyagawa [20]. In this section, we provide an application of the approximation for the kth nearest distance to locational analysis.



(a) Hospital (b) Police station (c) Post office

Figure 5: Facilities on the road network of Tsukuba

Suppose that facilities are closed independently and at random. Let p be the probability that each facility is open. Customers are assumed to be uniformly distributed and use their nearest open facility. Then the probability that customers use the second nearest facility, that is, the nearest facility is closed and the second nearest facility is open, is $(1-p)p$. In general, the probability that customers use the kth nearest facility is $(1-p)^{k-1}p$. Using this probability and the average kth nearest distance $E(R_k)$, we can express the average distance from customers to the nearest open facility $E(R)$ as

$$E(R) = p\sum_{k=1}^{\infty}(1-p)^{k-1}E(R_k). \tag{5.1}$$

Miyagawa et al. [21] and Miyagawa [19, 20] obtained the upper and lower bounds of $E(R)$ to find the optimal pattern of facilities that minimizes the average distance. For estimating the service level when some of the existing facilities are closed, the approximation for $E(R)$ would be helpful. Substituting the approximations (2.2) and (3.2) into (5.1) yields approximations for the average Euclidean and rectilinear distances $E(R^U), E(R^R)$ as

$$E(R^U) \approx \frac{p}{(1-p)\sqrt{\rho\pi}}\mathrm{Li}_{-\frac{1}{2}}(1-p), \tag{5.2}$$

$$E(R^R) \approx \frac{p}{(1-p)\sqrt{2\rho}}\mathrm{Li}_{-\frac{1}{2}}(1-p), \tag{5.3}$$

where $\mathrm{Li}_s(z) = \sum_{k=1}^{\infty} z^k/k^s$ is a polylogarithm. These approximations are depicted in Figure 7, where the density of facilities is $\rho = 1$. The upper and lower bounds of $E(R)$ of the grid pattern and $E(R)$ of the random pattern obtained in [19–21] are also shown in the figure. The approximation is greater than $E(R)$ of the grid and random patterns when p is

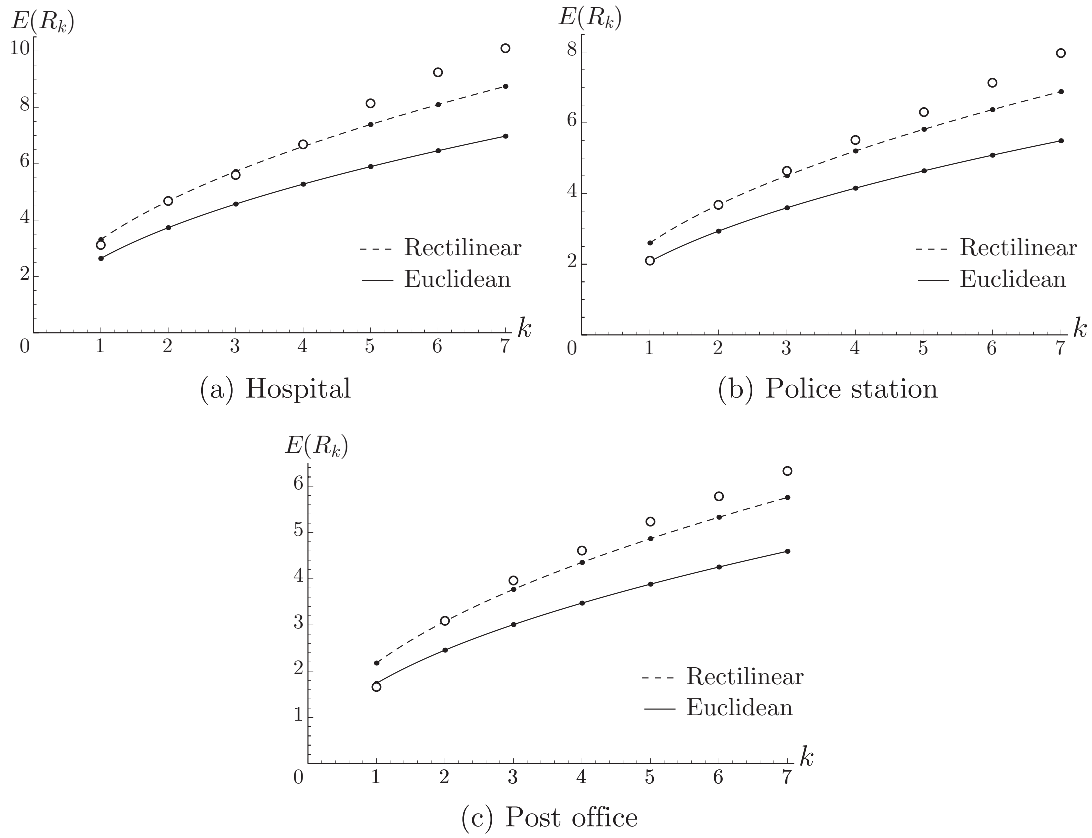


Figure 6: Average kth nearest network distance

large, because the approximations (2.2) and (3.2) overestimate the kth nearest distance for small k (see Figures 2 and 4).

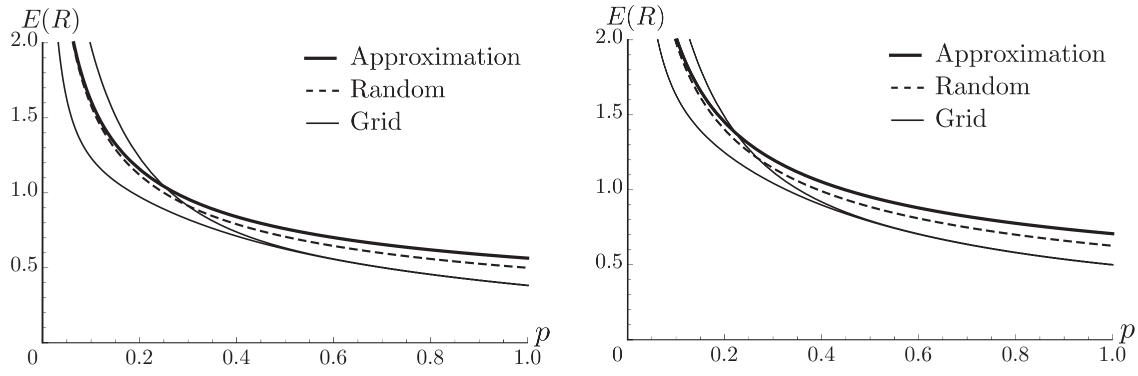
Suppose now that customers use up to the nth nearest facility. If all of the n facilities are closed, customers give up using facilities. The average distance is then rewritten as

$$E(R) = p \sum_{k=1}^{n} (1-p)^{k-1} E(R_k) + (1-p)^n A, \tag{5.4}$$

where A is a penalty for failing to use facilities. Substituting the approximations (2.2) and (3.2), we have approximations for the average distance $E(R)$ as shown in Figure 8. It can be seen that $E(R)$ increases with A and decreases with n. This implies that, if the penalty is large, customers should search for an open facility, even though the distance to the facility is great. The result makes sense intuitively, because customers in an emergency will certainly use a facility.

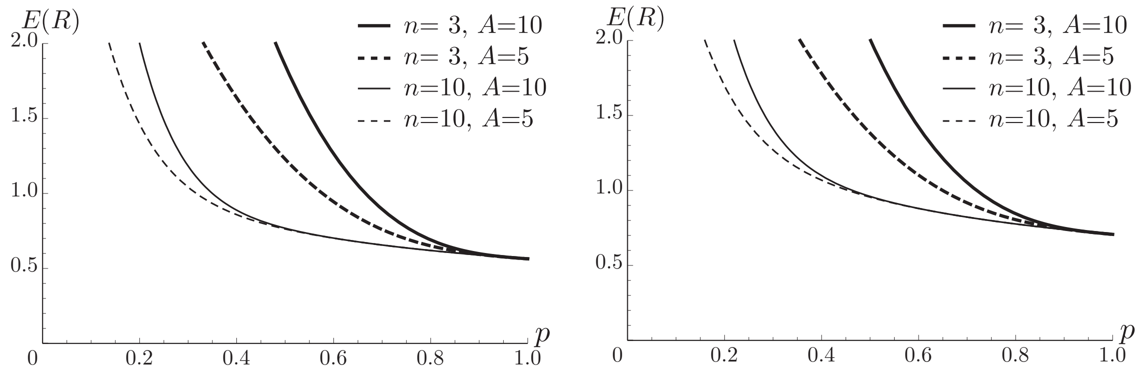
6. Conclusion

This paper has provided a simple approximation for the kth nearest distance. The approximation, which becomes more accurate as k increases, allows us to estimate higher order distances that have not previously been derived. The approximation demonstrates that the kth nearest distance is proportional to the square root of k and inversely proportional to the square root of the density of points. These relationships lead to a better understanding of fundamental characteristics of the kth nearest distance.



(a) Euclidean distance (b) Rectilinear distance

Figure 7: Average distance to the nearest open facility



(a) Euclidean distance (b) Rectilinear distance

Figure 8: Average distance to the nearest open facility

The approximation for the kth nearest distance will be useful for further facility location problems with closing of facilities. Using the kth nearest distance, we have obtained the average distance to the nearest open facility when some of the existing facilities are closed. This average distance gives an estimate for the service level of actual facility location. By comparing the average distances, we can evaluate the efficiency of actual location. In addition, the analytical expression for the average distance helps planners in estimating the number of facilities required to achieve a certain level of service. The estimated number of facilities can be used as an input in location models. The effect of a closing policy (e.g. seismic retrofitting) can also be assessed in terms of the reduction in the average distance.

Acknowledgements

I am grateful to two anonymous referees and the editor for their helpful comments and suggestions.

References

[1] N. Aras, M. Orbay, and I. Altinel: Efficient heuristics for the rectilinear distance capacitated multi-facility Weber problem. *Journal of the Operational Research Society*, **59** (2008), 64–79.

[2] O. Berman, D. Krass, and M. Menezes: Facility reliability issues in network p-median

problems: Strategic centralization and co-location effects. *Operations Research*, **55** (2007), 332–350.

[3] P. Clark and F. Evans: Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, **35** (1954), 85–90.

[4] N. Cressie: *Statistics for Spatial Data* (John Wiley & Sons, New York, 1993).

[5] M. Dacey: Two-dimensional random point patterns: A review and an interpretation. *Papers of the Regional Science Association*, **13** (1968), 41–55.

[6] Z. Drezner: Heuristic solution methods for two location problems with unreliable facilities. *Journal of the Operational Research Society*, **38** (1987), 509–514.

[7] R. Francis, F. McGinnes, and J. White: *Facility Layout and Location: An Analytical Approach* (Prentice Hall, Upper Saddle River, 1992).

[8] A. Getis: Second-order analysis of point patterns: The case of Chicago as a multi-center urban region. *The Professional Geographer*, **35** (1983), 73–80.

[9] F. He and R. Duncan: Density-dependent effects on tree survival in an old-growth Douglas fir forest. *Journal of Ecology*, **88** (2000), 676–688.

[10] P. Holgate: The distance from a random point to the nearest point of a closely packed lattice. *Biometrika*, **52** (1965), 261–263.

[11] P. Holgate: Some new tests of randomness. *Journal of Ecology*, **53** (1965), 261–266.

[12] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan: *Statistical Analysis and Modeling of Spatial Point Patterns* (John Wiley & Sons, Chichester, 2008).

[13] A. Jones, I. Langford, and G. Bentham: The application of K-function analysis to the geographical distribution of road traffic accident outcomes in Norfolk, England. *Social Science & Medicine*, **42** (1996), 879–885.

[14] T. Koshizuka: On the relation between the density of urban facilities and the distance to the nearest facility from a point in a given area. *Journal of the City Planning Institute of Japan*, **20** (1985), 85–90 (in Japanese).

[15] E. Krause: *Taxicab Geometry: An Adventure in Non-Euclidean Geometry* (Dover, New York, 1987).

[16] R. Larson: A hypercube queuing model for facility location and redistricting in urban emergency services. *Computers & Operations Research*, **1** (1974), 67–95.

[17] R. Larson and A. Odoni: *Urban Operations Research* (Prentice-Hall, Englewood Cliffs, 1981).

[18] V. Marianov and C. ReVelle: The queueing maximal availability location problem: A model for the siting of emergency vehicles. *European Journal of Operational Research*, **93** (1996), 110–120.

[19] M. Miyagawa: Analysis of facility location using ordered rectilinear distance in regular point patterns. *FORMA*, **23** (2008), 89–95.

[20] M. Miyagawa: Order distance in regular point patterns. *Geographical Analysis*, **41** (2009), 252–262.

[21] M. Miyagawa, Y. Ohsawa, and T. Koshizuka: Changes of travel distances by facility opening/closing and the robust regular location. *Transactions of the Operations Research Society of Japan*, **47** (2004), 1–24 (in Japanese).

[22] M. O'Kelly: Rectilinear minimax hub location problems. *Journal of Geographical Systems*, **11** (2009), 227–241.

[23] O. Persson: Distance methods: The use of distance measurements in the estimation of seedling density and open space frequency. *Studia Forestalia Suecica*, **15** (1964), 1–68.

[24] H. Pirkul: The uncapacitated facility location problem with primary and secondary facility requirements. *IIE Transactions*, **21** (1989), 337–348.

[25] B. Ripley: The second-order analysis of stationary point processes. *Journal of Applied Probability*, **13** (1976), 255–266.

[26] L. Snyder and M. Daskin: Reliability models for facility location: The expected failure cost case. *Transportation Science*, **39** (2005), 400–416.

[27] P. Sorensen and R. Church: Integrating expected coverage and local reliability for emergency medical services location problems. *Socio-Economic Planning Sciences*, **44** (2010), 8–18.

[28] H. Thompson: Distribution of distance to nth neighbour in a population of randomly distributed individuals. *Ecology*, **37** (1956), 391–394.

[29] R. Vaughan: *Urban Spatial Traffic Patterns* (Pion, London, 1987).

[30] J. Weaver and R. Church: A median location model with nonclosest facility service. *Transportation Science*, **19** (1985), 58–74.

Masashi Miyagawa
Department of Ecosocial System Engineering
University of Yamanashi
4-3-11 Takeda, Kofu
Yamanashi 400-8511, Japan
E-mail: mmiyagawa@yamanashi.ac.jp