Science on Form: Proceedings of the First International Symposium for Science on Form,
General Editor: S. Ishizaka, Editors: Y. Kato, R. Takaki, and J. Toriwaki, 491-498.

Derivation of the Probability Density Function of Distance and Its Applications to Urban Facility Planning

*Takeshi Koshizuka and **Yoshiaki Ohsawa

**Institute of Socio-Economic Planning, University of Tsukuba*
***Doctoral Degree Programs in Socio-Economic Planning, University of Tsukuba*

Keywords: Voronoi Diagram, Nearest-neighbour distance

The objective of this paper is to gauge the level of service of facilities, such as elementary and junior high schools as well as public offices, by the use of probability density function of distance between residents' locations and the facilities. The first part shows the derivation of the probability density function of distance. The second part contains the special derivation of the probability density function of the nearest-neighbour distance using Voronoi diagram. The third and last part demonstrates its application to location planning of elementary schools in Ohmiya City, about 30 kilometers north of Tokyo Metropolitan Area.

Since the work of Clark and Evans (1954), a considerable number of studies on point pattern with respect to distance have been undertaken by Thompson (1956), Pielou (1969), Okudaira and Koshizuka, et. al. (1979), Ripley (1981), and several others. However, little analysis has been advanced because we cannot analytically derive the probability density function (pdf) of distance between a sample point and facility location points without imposing assumptions on the point pattern as perfectly regular or random.

This paper is organized as follows. In the next Section the method for deriving the pdf of distance in any point pattern by a "right-angled triangulation procedure" is presented. In Section 3, the pdf of nearest-neighbour distance using Voronoi diagram is derived. In addition, if the points are considered to be facilities, the distance to each facility can be regarded as a measure of the level of service of the facility. The application of the method to elementary school planning in Ohmiya City in Japan is described in the final Section.

2. Derivation of the Probability Density Function of Distance

Consider a region where the following are given:
(a) the location of N facilities,
(b) the pdf of population in this region, and
(c) the location of vertices which form the boundary of the catchment area.

The pdf and moments of distance between these residential locations and the facilities in this region will be derived based on the above assumptions.

2.1. Derivation of the General Form of Probability Density Function

Without loss of generality, using Cartesian coordinates, the pdf of population in the region can be represented by

$$f(x,y) = \sum_{i=1}^{N} \lambda_i(x,y) \tag{2.1}$$

where

$\lambda_i(x,y)$ = the population density in the i-th district denoted by D_i

such that

$$\lambda_i(x,y) \geq 0 \quad \text{and} \quad \sum_{i=1}^{N} \iint_{D_i} \lambda_i(x,y)dxdy = 1.$$

For each district, using a facility location (x_{i0},y_{i0}), the variable (X,Y) can be transformed into (R,Θ), where $R = \sqrt{(x-x_{i0})^2+(y-y_{i0})^2}$, $\Theta = \arctan[(y-y_{i0})/(x-x_{i0})]$, hence, equation (2.1) becomes

$$f(r,\theta) = r \sum_{i=1}^{N} \rho_i(r,\theta) \tag{2.2}$$

where

$$\rho_i(r,\theta) = \lambda_i(r\cos\theta + x_{i0}\ ,\ r\sin\theta + y_{i0}).$$

Integrating equation (2.2) over θ, the marginal distribution of randam variable R, that is, the pdf of distance in this region can be obtained as follows:

$$f(r) = r\left[\sum_{i=1}^{N}\sum_{j=1}^{M_i(r)} \int_{\theta_{ij1}(r)}^{\theta_{ij2}(r)} \rho_i(r,\theta)d\theta\right], \tag{2.3}$$

where

$M_i(r)$ = the number of the integration domains in the i-th district for radius r as illustrated in Figure 2.1,

$\theta_{ij1}(r)$ = the appropriate lower limit of the j-th domain in the i-th district, and

$\theta_{ij2}(r)$ = the appropriate upper limit of the j-th domain in the i-th district

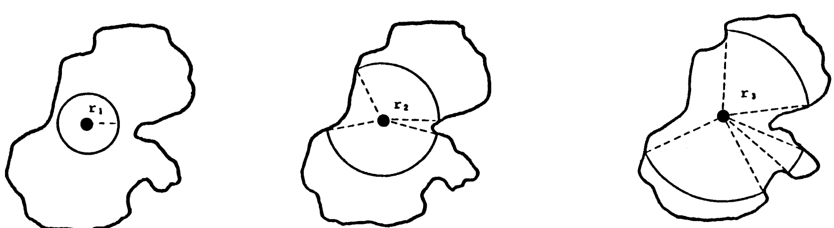


Figure 2.1 Number of the integration domains. (a) $m(r_1) = 1$, (b) $m(r_2) = 2$, (c) $m(r_3) = 3$. ● indicates the location of facility.

Equation (2.3) expresses the general form of the probability density function of distance.

Suppose, however, in each district the population distribution depends only on R such that:

$$\rho_i(r,\theta) = \rho_i(r).$$

Equation (2.3) can be reduced to

$$f(r)= \sum_{i=1}^{N} \rho_i(r)\, L_i(r), \qquad (2.4)$$

where

$$L_i(r) = r \sum_{i=1}^{N} (\theta_{ij2}(r)-\theta_{ij1}(r)),$$ i.e., the length of circular arc in the i-th district for radius r.

The right-angled triangulation procedure which can systematically calculate $L_i(r)$ in equation (2.4) by the use of computer will be developed in the next Section.

2.2. Right-angled Triangulation Procedure

The calculaton of $L_i(r)$ seems to be a complicated task. However, it can be calculated systematically by employing the "right-angled triangulation procedure", a procedure which is analogous to the combinatorial principle of inclusion-exclusion, and is described below.

Any district can generally be bisected into a number of triangles, each of which is formed by two vertices and one facility point as shown in Figure 2.2(a) and (b). The sum of all the triangles minus the triangles beyond the zone boundary would give the coverage of this district. Any triangle can generally be expressed by two right-angled triangles, each of which is formed by one facility point, one vertex, and one foot of the perpendicular of the boundary edge from the facility. This expression is of two variations - one is the union of the two right-angled triangles; the other is the difference between them as illustrated in Figure 2.3. Hence, any district in the region can be expressed by an aggregate of right-angled triangles.

In simple triangulation, the determination of signs of triangles is a cumbersome procedure. The following procedure has been devised to systematically attach the signs to each triangle:

(1) Number the vertices and their corresponding triangles in counterclockwise direction according to the boundary from a vertex denoted by Q (see Figure 2.2(b) and (c)).

(2) Calculate the angle $\alpha(-\pi \leq \alpha \leq \pi)$ between the line segment joining the j-th vertex to the facility point and the one joining the j+1-th vertex (if j = n ,the first vertex instead of the n+1-th vertex because of cyclic order) to the facility point (see Figure 2.2(c)).

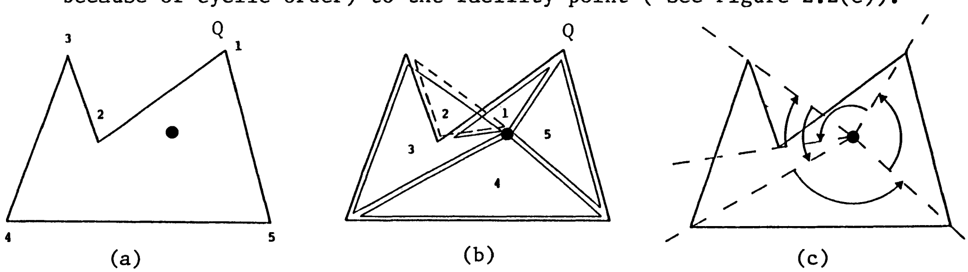


Figure 2.2 Triangulation

(3) Define the sign variable of the j-th triangle in the i-th district as follows:

$$\delta_{ij} \begin{cases} = 1 & (\text{if } \alpha \geqq 0) \\ = -1 & (\text{otherwise}). \end{cases}$$

In the right-angled triangulation, it is easy to determine the sign of right-angled triangles by the following procedure (see Figure 2.3).

(1) Label the vertex farther from the facility point as A and the other as B.

(2) Define the sign variable of two right-angled triangles of the j-th triangle in i-th district as follows:

$$\xi_{ij1} = 1$$

$$\xi_{ij2} \begin{cases} = 1 & (\text{if } OA^2 \leqq OB^2 + AB^2) \\ = -1 & (\text{otherwise}) \end{cases} .$$

It is therefore recognized that any district can be systematically represented by an aggregate of right-angled triangles. Introducing the two sign variables defined above, the length of arc in the i-th district , $L_i(r)$, can be expressed as follows:

$$L_i(r) = \sum_{j=1}^{M_i} \delta_{ij} \sum_{k=1}^{2} \xi_{ijk} L_{ijk}(r) \tag{2.5}$$

where

M_i = the number of triangles in the i-th district,

$L_{ijk}(r)$= the length of a circular arc of the k-th right-angled triangle in the j-th triangle in the i-th district.

Now, $L_{ijk}(r)$ can be calculated easily. From the locations of the two boundary vertices and the facility point, the length of two edges denoted by a and b ($a \geqq b$) can be calculated. $L_{ijk}(r)$ can be represented by (see Figure 2.4):

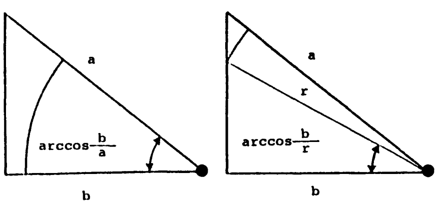
$$L_{ijk}(r) \begin{cases} = r \arccos(b/a) & (\text{if } 0 \leqq r \leqq b) \\ = r\,[\arccos(b/a) - \arccos(b/r)] & (\text{if } b \leqq r \leqq a) \\ = 0 & (\text{otherwise}). \end{cases} \tag{2.6}$$

Hence, from equation (2.4),(2.5) and (2.6), the pdf of distance to the facility in the region can be specified as follows:

$$f(r) = \sum_{i=1}^{N} \rho_i(r) \sum_{j=1}^{M_i} \delta_{ij} \sum_{k=1}^{2} \xi_{ijk} L_{ijk}(r). \tag{2.7}$$

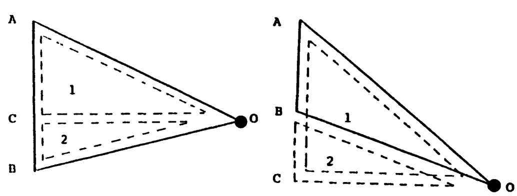
If the population density in each district is uniform such that $\rho_i(r) = \rho_i$, equation (2.7) can be reduced to

$$f(r) = \sum_{i=1}^{N} \rho_i \sum_{j=1}^{M_i} \delta_{ij} \sum_{k=1}^{2} \xi_{ijk} L_{ijk}(r). \tag{2.8}$$



(a) (b)

Figure 2.3 Right-angled triangulation (a) union, (b) difference.



(a) (b)

Figure 2.4 A circular arc in the right-angled triangle, (a) $0 \leqslant r \leqslant b$, (b) $b \leqslant r \leqslant a$.

Similarly if the population density in the whole region is uniform such that $\rho_i = 1/S$ where S is the area of the whole region, equation (2.7)

reduces to

$$f(r)= \sum_{i=1}^{N} \sum_{j=1}^{M_i} \delta_{ij} \sum_{k=1}^{2} \xi_{ijk} L_{ijk}(r)/S. \qquad (2.9)$$

2.3. Cumulative Distribution Function, Expectation, and Variance

The cumulative distribution function (cdf) of R for equations (2.8) and (2.9) can be obtained by integrating equation (2.6).

$$\int_0^r L_{ijk}(u)du$$

$$\begin{cases} = r^2 \arccos(b/a)/2 & (\text{if } 0 \leq r \leq b) \\ = r^2 [\arccos(b/a)-\arccos(b/r)]/2+b\sqrt{r^2-b^2}/2 & (\text{if } b \leq r \leq a) \\ = b\sqrt{a^2-b^2}/2 & (\text{if } a \leq r). \end{cases}$$

The expectation and variance of R can be derived from equation (2.8) and (2.9) respectively:

$$\int r\, L_{ijk}(r)dr = b\,\{a\sqrt{a^2-b^2}+b^2 \log[(a+\sqrt{a^2-b^2})/b]\}/6, \qquad (2.10)$$

$$\int r^2 L_{ijk}(r)dr = b\,\sqrt{a^2-b^2}(a^2+2b^2). \qquad (2.11)$$

Similarly, by specifying the function $\rho_i(r)$, the cdf, the expectation and variance using equation (2.7) can also be derived.

3. Probability Density Function of Nearest-neighbour Distance

Consider a region with given:
(a) location of N facilities, and
(b) pdf of population in this region.

Assume that the inhabitants always go to the nearest-neighbour facility whenever possible. Based on these assumptions, the pdf of nearest-neighbour distance will be derived.

3.1. Voronoi Diagram

The Voronoi diagram also known as the Dirichlet tesselation or as the Thiessen polygons is described as follows: Let $x_1, x_2, \cdots, x_n$ be Cartesian coordinates of N points, then Voronoi polygon $V(x_i)$ for point x_i is defined by

$$V(x_i) = \{x \mid d(x,x_i) \leq d(x,x_j)\},$$

where $d(\cdot,\cdot)$ = the Euclidean distance. Voronoi Diagram consists of the boundaries of the Voronoi polygons. This diagram forms a non-overlapping covering of the plane.

3.2. Probability Density Function of Nearest-neighbour Distance

Nearest-neighbour distance is defined as the distance between a sample point and its nearest neighbour point. Clark and Evans (1952) obtained this distribution for a random point pattern

By using Monte-Carlo simulation, Ripley (1980) obtained the approximate pdf's of nearest-neighbour distance on several typical patterns of actual data. However, utilizing the Voronoi diagram and the right-angled triangulation procedure, this distribution can be analytically derived.

Essentially, a Voronoi diagram is a partition of the plane into N Voronoi polygons, each of which is associated with a given generator.

The Voronoi polygon with a generator is the locus of points closer to that point other than any given point. The Voronoi diagram can be considered to represent catchment areas where the inhabitants tend to go to the nearest-neighbour point whenever possible. Therefore, the pdf of nearest-neighbour distance can be derived by considering the Voronoi diagram as the cathchment areas. The Voronoi diagram can be viewed as the optimal allocation of a region in terms of distance.

4. A Case Study for the Planning of Elementary School in Ohmiya City

The above methodology is applied to planning elementary schools in Ohmiya City which is about 30 kilometers north of Tokyo Metropolitan Area. In order to derive the pdf of distance, the road distance is replaced by the Euclidean distance for the sake of simplicity. This simplification is however not so restrictive because the road distance is empirically proven to be proportional to the Euclidean distance [Koshizuka and Kobayashi (1984)].

Figure 4.1 shows the population distribution map of Ohmiya City in 1975, each dot representing approximately 25 persons. Figure 4.2 shows the existing 35 elementary schools in Ohmiya City, and the actual school districts and the Voronoi diagram which is constructed by the locations of elementary schools as generators.

Assuming that the population density is uniform, the pdf of distance in the school districts and the pdf of nearest-neighbour distance can be obtained from equation (2.8). The total population in 1,854 census districts (taken from the 1975 Population Census of Japan) was used to estimate the pupil population density in each school district or Voronoi polygon. The census districts were delineated to comprise about 50 households as a basic spatial unit in the Population Census in Japan. The population in each district or polygon is defined to be the sum of population of the census districts if their centroids fall within the district or polygon. This operation can be undertaken by the

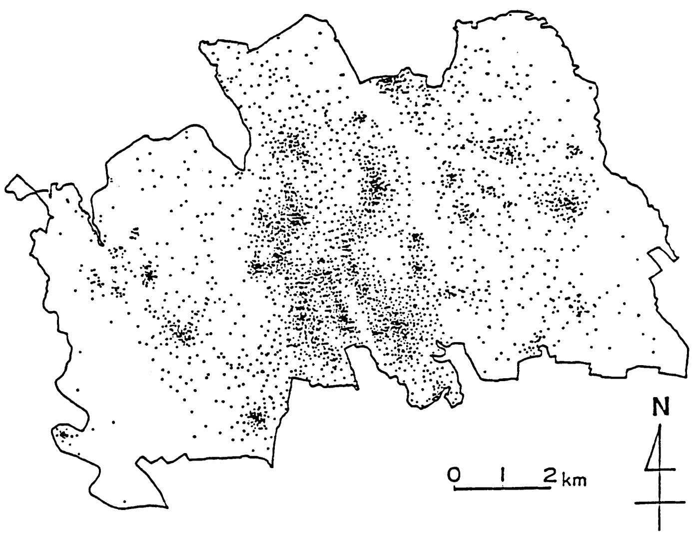


Figure 4.1 Population distribution map in 1975 (each dot represents approximately 25 persons).

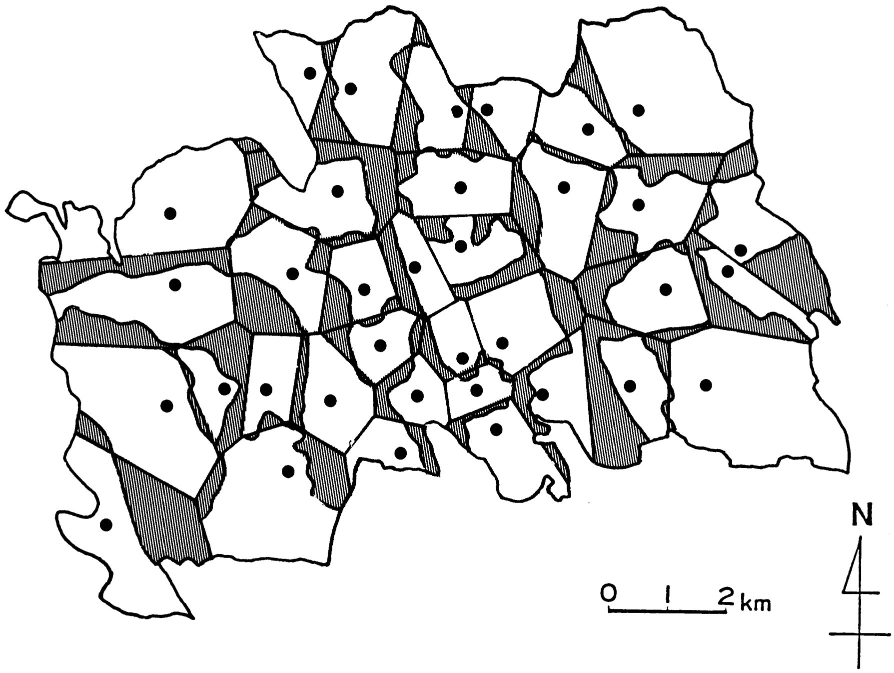


Figure 4.2 Difference area between actual school districts and Voronoi diagram. ● indicates the existing elementary school in Ohmiya City.

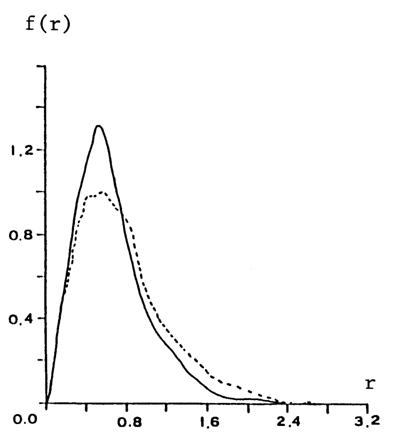


Figure 4.3 P.d.f. of distance of school trip (dotted line) and that of the nearest-neighbour distance (solid line).

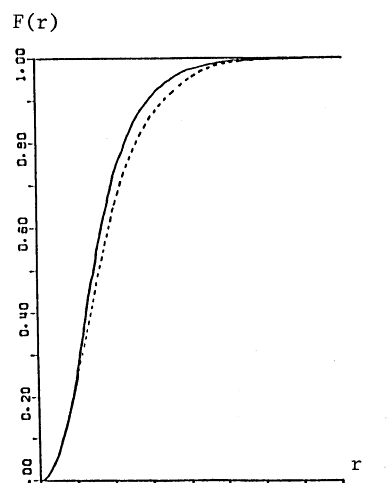


Figure 4.4 C.d.f. of distance of school trip (dotted line) and that of the nearest-neighbour distance (solid line).

Table 4.1 Expectations and standard deviations (km).

	expectation	standard deviation
nearest-neighbour distance	0.646	0.387
distance of school trip	0.736	0.443

point-inclusion algorithm. The efficiency of these algorithms are compared in Edahiro, Kokuba and Asano (1984).

The pdf's and cdf's of distance in existing school districts and in Voronoi diagram using these population data and equation (2.8) are shown in Figure 4.4 and 4.5. Their computed moments using equation (2.10) and (2.11) are listed in Table 4.1. From these, we can conclude that if each catchment area is shifted from the existing school districts to the Voronoi diagram without changing the locations of elementary schools, the distance of school trip can be reduced by approximately 90 meters per person. This reduction is due to the fact that the area of the school districts and Voronoi polygons do not coincide as shown in the shaded areas in Figure 4.3. Since the difference is only 90 meters per person, we can say that these school districts are nearly optimal.

5. Conclusion

In this paper, the derivation of the pdf of distance to a facility has been studied. In the application of this method in Ohmiya City, results showed that the distance of school trip can be reduced by approximately 90 meters per person by transforming the configuration of the existing school district to Voronoi polygons.

References

Clark P.J. and F.C. Evans [1954]: Distance to Nearest Neighbour as a Measure of Spatial Relationships in Populations. Ecology, 35, pp.445-453.

Edahiro M., I. Kokubo and T.Asano [1984]: A New Point Location Algorithm and Its Practical Efficiency---Comparison with Existing Algorithms. ACM Trans. on Graphics, 3, pp.86-109.

Getis A. and B. Boots [1978]: Models of Spatial Processes. Cambridge University Press, London.

Koshizuka T. and Y. Ohsawa [1983]: A Probability Density Function of Distance in Urban Facility Planning (in Japanese). Papers of the Annual Conference of the City Planning Institute of Japan, 18, pp.25-30.

Ohya T., M. Iri and K. Murota [1984]: Improvements of the Incremental Method for the Voronoi Diagram with Computational Comparison of Various Algorithms. J. Oper. Res. Soc. Japan, 27, pp.306-337.

Okudaira T., T. Koshizuka, et. al. [1979]: Regional Research (NO.1) (in Japanese). Communication of the Operations Research Society of Japan, 24(3), pp.152-159.

Pielou E.C. [1969]: An Intoduction to Mathematical Ecology. John Wiley, New York.

Ripley B.D. [1981]: Spatial Statistics. John Wiley, New York.

Shamos M.I. [1978]: Computational Geometry. Ph.D. Thesis, Yale University.

Shamos M.I. and D. Hoey [1975]: Closest-Point Problems. Proceedings of the 16th Annual IEEE Symposium on Fundations of Computer Science, pp.151-162.

Thompson H.R. [1956]: Distribution of Distance to N-th Neighbour in a Population of Randomly Distributed Individuals. Ecology, 37, pp.391-394.