Distance distributions

Ten Raa, M.H.

Published in:
Geographical Analysis: An International Journal of Theoretical Geography

Publication date:
1983

Distance Distributions *by Thijs ten Raa*

Distance distributions are important in the analysis of spatial phenomena. They arise in many problems in distribution management as noted by Eilon, Watson-Gandy, and Christofides (1971, p. 151), in shape and migration measurements by Taylor (1971, pp. 42–51), and also in a study of commuting by Verster (1979, pp. 27–29). This paper determines distance distributions for arbitrary domains and distances.

Section 1 considers various distance distributions put forward in Kuiper and Paelinck (1982). The so-called contact frequency is criticized and corrected. The distance frequency is borrowed without modification. Section 2 calculates the moments of the distributions. Section 3 applies the results on the square with the Manhattan distance for comparison with Eilon, Watson-Gandy, and Christofides (1971) and Kuiper and Paelinck (1982). Section 4 determines the distributions themselves. Section 5 concludes the paper.

1. Definitions

We consider a measurable domain Ω in a space equipped with a Minkowski distance $\|.-.\|_p$, $1 \leq p \leq \infty$. . This distance is undefined for $p < 1$. $p = 1$ represents the Manhattan distance. $p = 2$ represents the Euclidean distance. $p > 2$ represents the even shorter distances. Since Ω must merely be measurable, we need not decide what general shape a zone under consideration most closely resembles. Contrast this with Kuiper and Paelinck (1982) who confine themselves to the circle, square, and rectangle. One still has to decide upon a value of p. This is a matter of empirical geography. For the real world one has roughly $p = 1.5$; see Love and Morris (1979, pp. 136–37) who also conclude that the Minkowski distance is superior, at least for urban domains.

The first distribution we shall deal with is the so-called contact frequency. Kuiper and Paelinck (1982) measure the amount of contact of x with points on distance l as

$$\left(\int_{\|x-y\|_p = l,\, y \in \Omega} ds/c_p l \right) \int_{\|x-y\|_p = l,\, y \in \Omega} ds \,,$$

where ds is a differential element along the arc $\|x - y\|_p = l$, $y \in \Omega$, that is, the set of all points y which lie at distance l from x but within our domain Ω; and c_p is the circumference of the unit circle under the p-distance. (E.g., $c_1 = 4\sqrt{2}$, $c_2 = 2\Pi$, and $c_\infty = 8$ of the circles in the plane.) When, for given l, the length of the arc is doubled, for example, by alternative choice of x, then the amount of contact is quadrupled according to Kuiper and Paelinck. This is counterintuitive. Their measure should be corrected. The amount of contact of x with points on distance

Thijs ten Raa is assistant professor of theoretical spatial economics, Erasmus University Rotterdam.

Geographical Analysis, Vol. 15, No. 2 (April 1983)

l should be measured by

$$\int_{\|x-y\|_p=l,\, y\in\Omega} ds \,,$$

that is, the length of the arc itself. When the full circle about x with radius l belongs to Ω, then the definitions coincide. Otherwise Kuiper and Paelinck tend to underestimate the amount of contact as they enter the arc fraction twice into the measure. In sum, we define the *contact frequency* cf_p by

$$cf_p(l) = \int_\Omega \left(\int_{\|x-y\|_p=l,\, y\in\Omega} ds \right) dx \,.$$

The second distribution will be taken from Kuiper and Paelinck (1982) directly. The *distance frequency* df_p is defined by

$$df_p(l) = \int_\Omega \left(\int_{\|x-y\|_p=l,\, y\in\Omega} ds/c_p l \right) dx \,.$$

Their road area frequency is not considered here.

Besides the circumference c_p of the unit circle under the p-distance it is useful to define its area d_p. (E.g., $d_1 = 2$, $d_2 = \Pi$, and $d_\infty = 4$ in R^2.) The *nth moment* of cf_p is defined by

$$cf_p^{(n)} = E(l^n|cf_p) = \int_0^\infty l^n cf_p(l) dl$$

and similarly for df_p. Useful shorthand is $I_p^{(n)}(\Omega) = \int_\Omega \int_\Omega \|x - y\|_p^n dx dy$, that is, the integral of the nth power of the p-distance over all pairs in Ω.

2. *Moments*

Our first theorem offers explicit formulas for the moments of our distributions.

THEOREM 1

$$cf_p^{(n)} = c_p d_p^{-1} I_p^{(n)}(\Omega) \text{ and } df_p^{(n)} = d_p^{-1} I_p^{(n-1)}(\Omega) \,.$$

Proof

$$cf_p^{(n)} = \int_0^\infty l^n cf_p(l) dl = \int_0^\infty l^n \left[\int_\Omega \left(\int_{\|x-y\|_p=l,\, y\in\Omega} ds \right) dx \right] dl$$

$$= \int_0^\infty \int_\Omega \left(\int_{\|x-y\|_p=l,\, y\in\Omega} l^n ds \right) dx \, dl \,.$$

By Fubini's theorem,

$$cf_p^{(n)} = \int_{\|x-y\|_p = l, y\in\Omega} \int_0^\infty \int_\Omega l^n ds dl dx \ .$$

By change of variable $(s,l) \leftrightarrow y$, which has Jacobian c_p/d_p,

$$cf_p^{(n)} = \int_\Omega \int_\Omega \|x - y\|_p^n c_p d_p^{-1} \, dydx = c_p d_p^{-1} I_p^{(n)}(\Omega) \ .$$

In the same way,

$$df_p^{(n)} = \int_0^\infty l^n \left[\int_\Omega \left(\int_{\|x-y\|_p = l, y\in\Omega} ds/c_p l \right) dx \right] dl \ .$$

By Fubini's theorem,

$$df_p^{(n)} = \int_{\|x-y\|_p = l, y\in\Omega} \int_0^\infty \int_\Omega l^{n-1}/c_p ds dl dx \ .$$

By change of variable,

$$df_p^{(n)} = d_p^{-1} \, I_p^{(n-1)} (\Omega) \ .$$

COROLLARIES

1. $cf_p^{(n)} = c_p \, df_p^{(n+1)}$.
2. *Average distance equals* $df_p^{(1)}/df_p^{(0)} = cf_p^{(0)}/(c_p \, df_p^{(0)})$, *that is, total contacts over the product of the unit-circle circumference and the total distance.*
3. *Average contact distance equals* $cf_p^{(1)}/cf_p^{(0)} = I_p^{(1)}(\Omega)/I_p^{(0)}(\Omega) = \int_\Omega \int_\Omega \|x - y\|_p \, dxdy/ (\int_\Omega dx . \int_\Omega dy)$, *that is, the expected distance between two points chosen at random in the domain. Note that this expectation is based on the contact frequency and not on the distance frequency as (erroneously) suggested by Kuiper and Paelinck (1982).*

3. *Application*

To compare our formulas with some of Eilon, Watson-Gandy, and Christofides (1971) and Kuiper and Paelinck (1982), we zero in on their square, $\Omega = [0,a]^2$, and Manhattan distance, $p = 1$.

LEMMA

$I_1^{(n)}([0,a]^2)$ *equals* $(4/3)(4\log 2 - 1)a^{n+4}$ for $n = -1$ *and* $.8(2^{n+3} - n - 5)(n!/(n + 4)!)a^{n+4}$ for $n \geq 0$.

Proof is by change of variables $(x,y) \leftrightarrow (u,v) = (x - y, x + y)$—which has Jacobian 1/2—and calculus, which is omitted here.

By theorem 1 and the lemma, the average contact distance equals

$$
\begin{aligned}
cf_1^{(1)}/cf_1^{(0)} &= I_1^{(1)}([0,a]^2)/I_1^{(0)}([0,a]^2) \\
&= 8(2^4 - 1 - 5)\frac{1!}{5!}a^5 \Big/ \left[8\left(2^3 - 0 - 5\right)\frac{0!}{4!}a^4\right] \\
&= \frac{16-6}{5!}\frac{4!}{8-5}a = \frac{10}{5.3}a = 0.667\, a\ .
\end{aligned}
$$

This agrees with the result of Eilon, Watson-Gandy, and Christofides, but exceeds that of Kuiper and Paelinck as was to be expected because of our correction of their contact-frequency definition. By theorem 1 and the lemma, the average distance equals

$$
\begin{aligned}
df_1^{(1)}/df_1^{(0)} &= I_1^{(0)}([0,a]^2)/I_1^{(-1)}([0,a]^2) \\
&= 8(2^3 - 0 - 5)\frac{0!}{4!}a^4/\left[\frac{4}{3}(4\log 2 - 1)a^3\right] \\
&= a/\left[\frac{4}{3}(4\log 2 - 1)\right] = 0.423\, a\ .
\end{aligned}
$$

This is the same as Kuiper and Paelinck's result.

4. Determination

Since we have precise expressions for the moments of the distributions rather than inaccurate statistics like higher-order sample moments, it is promising to derive the distributions from the moments calculated in the first theorem.

Our second theorem determines the distributions themselves. It hinges upon the boundedness of geographical domains. Hitherto, according to Taylor (1971, p. 40), this fact has been viewed as a source of the peculiarity of geographical distributions. But now the boundedness in geography proves to be a treasure trove.

THEOREM 2. *If Ω is bounded, then $cf_p^{(n)}$ and $df_p^{(n)}$ are finite and uniquely determine cf_p and df_p, respectively.*

Proof. The finiteness is obvious. When Ω is bounded, the ranges of cf_p and df_p are finite. By corollary 4.22(c) of Kendall and Stuart (1969), the moments uniquely determine each of the distributions.

We shall now attempt to find cf_p and df_p explicitly. We know their moments, that is, in products with the elementary polynomials, x^n. We want their values at l, that is, inproducts with the unit impulse concentrated at l, δ_l. It is natural to express the unit impulse in terms of polynomials, like

$$\delta_l(x) = \sum_0^{\infty} a_n x^n\ .$$

Then

$$cf_p(l) = \int_0^{\infty} \delta_l(x) cf_p(x) dx$$

$$= \int_0^{\infty} \sum_0^{\infty} a_n x^n cf_p(x) dx$$

$$= \sum_0^{\infty} a_n \int_0^{\infty} x^n cf_p(x) dx$$

$$= \sum_0^{\infty} a_n cf_p^{(n)}$$

$$= c_p d_p^{-1} \sum_0^{\infty} a_n I_p^{(n)} (\Omega)$$

by theorem 1 and similarly for df_p. Here $\int_0^{\infty} \Sigma_0^{\infty} = \Sigma_0^{\infty} \int_0^{\infty}$ and is finite by Lebesgue's theorem provided that $\Sigma_0^n a_k l^k cf_p(l)$ are dominated by a summable function. I believe this condition is fulfilled for the same reason that cf_p is uniquely determined by its moments (theorem 2). In short, we want formal moment expansions of the distributions and we expect these expansions to be material, for the moments completely determine the distributions.

To facilitate the determination of the coefficients in the expansions we shall employ orthogonal polynomials rather than the elementary ones. Then coefficients are simply inproducts with the normed polynomials. Thus, let $P_n(x) = \Sigma_{k=0}^n a_k^{(n)} x^k$ be the Legendre polynomials (i.e., $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = 1/2(3x^2 - 1), \ldots$). Then

$$\int_{-1}^{1} P_m(x) P_n(x) dx = \frac{2}{2n + 1}$$

if $m = n$ and zero otherwise. And, letting $\Delta_p = \sup_{x,y \epsilon \Omega} \|x - y\|_p$,

$$\delta_l(x) = \sum_{n=0}^{\infty} c_n^{(l)} P_n(x/\Delta_p)$$

with

$$c_n^{(l)} = \int_{-\Delta_p}^{\Delta_p} \delta_l(x) P_n(x/\Delta_p) dx \Big/ \int_{-\Delta_p}^{\Delta_p} P_n(x/\Delta_p) P_n(x/\Delta_p) dx$$

$$= P_n(l/\Delta_p) \Big/ \frac{2\Delta_p}{2n + 1} = \frac{n + (1/2)}{\Delta_p} P_n(l/\Delta_p) .$$

It follows that

$$cf_p(l) = \int_0^\infty \delta_l(x) cf_p(x) dx$$

$$= \int_0^\infty \sum_{n=0}^{\infty} \frac{n + (1/2)}{\Delta_p} P_n(l/\Delta_p) P_n(x/\Delta_p) cf_p(x) dx$$

$$= \sum_{n=0}^{\infty} \frac{n + (1/2)}{\Delta_p} P_n(l/\Delta_p) \int_0^\infty \sum_{k=0}^{n} a_k^{(n)} (x/\Delta_p)^k cf_p(x) dx$$

$$= \sum_{n=0}^{\infty} \frac{n + (1/2)}{\Delta_p} P_n(l/\Delta_p) \sum_{k=0}^{n} a_k^{(n)} cf_p^{(k)}/\Delta_p^k .$$

Invoking theorem 1 we obtain

$$cf_p(l) = c_p d_p^{-1} \Delta_p^{-1} \sum_{n=0}^{\infty} \left(n + \frac{1}{2}\right) P_n(l/\Delta_p) \sum_{k=0}^{n} a_k^{(n)} I_p^{(k)} (\Omega)/\Delta_p^k .$$

And in the same way,

$$df_p(l) = d_p^{-1} \Delta_p^{-1} \sum_{n=0}^{\infty} \left(n + \frac{1}{2}\right) P_n(l/\Delta_p) \sum_{k=0}^{n} a_k^{(n)} I_p^{(k-1)} (\Omega)/\Delta_p^k .$$

These are the desired expressions for cf_p and df_p.

5. *Conclusion*

Distance distributions for arbitrary domains and distances can be determined through their moments. Until now the distributions were determined by direct computations in very special cases such as the square with a Manhattan distance. The moments, however, can be calculated in general and then used to determine the distributions themselves.

Sometimes, as in shape measurements, distance distributions are merely computed to obtain so-called moment measures. In such situations the direct calculation of the moments developed in this paper renders a shortcut.

LITERATURE CITED

Eilon, S., C. D. T. Watson-Gandy, and N. Christofides (1971). *Distribution Management.* London: Griffin.

Kendall, M. G., and A. Stuart (1969). "Distribution Theory." Volume 1 of *The Advanced Theory of Statistics.* London: Griffin.

Kuiper, J. H., and J. H. P. Paelinck (1982). "Frequency Distributions of Distances and Related Concepts." *Geographical Analysis,* 14, 253–59.

Love, R. F., and J. G. Morris (1979). "Mathematical Models of Road Travel Distances." *Management Science,* 25, 130–39.

Taylor, P. J. (1971). "Distances within Shapes: An Introduction to a Family of Finite Frequency Distributions." *Geografiska Annaler,* 53B, 40–53.

Verster, A. C. P. (1979). *Ruimtelijk Gedrag van Werkers in West-Nederland.* Projectbureau Integrale Verkeers- en Vervoerstudies, Netherlands Ministry of Transportation.