UDC 551.509.3:519.21

Distance Distributions for Randomly Distributed Data[1]

J. J. STEPHENS—*Florida State University, Tallahassee, Fla.*

ABSTRACT—Probability density distributions are derived and verified for the distance from an arbitrary point to stations randomly distributed in the plane. These distributions are essential to the analysis of the filtering properties of objective analysis schemes.

1. INTRODUCTION

Arithmetic operations on data arrays derived from space- or time-sampling of continuous fields lead to spectral modification of the signal. The alteration for linear operations on regularly spaced data can be determined exactly with a discrete transform approach. However, even with linear operators acting on a band-limited signal, the analysis for irregularly distributed data is usually complicated by the nonstationary sampling pattern over the domain. Only an average response can be obtained for nonstationary patterns, and that is subject to modeling approximations to the distribution of data. Analysis leading to the appropriate power transfer function requires a characterization of the probability distribution of data locations from a randomly selected point.

The modeling approximation used by Stephens and Stitt (1970) and Stephens and Polan (1971) presumed a random distribution of observing sites in the plane. However, the distance distribution used in determining the power transfer function was fitted empirically. While valid, their results are not readily utilized because the distribution form was incorrect. A theoretical formulation of the distribution is given and verified here.

2. DISTANCE DISTRIBUTIONS

The modeling approximation used here assumes that the observation sites are randomly distributed in the plane. The number of stations to be found in an area then follows the Poisson distribution (Parzen 1960). If the average station density, $\bar{\eta}$, is

$$\bar{\eta}=\lim_{A\to\infty}\frac{N}{A} \tag{1}$$

where N is the number of stations in the area, A, then the probability that exactly n stations are in A is

$$P[N(A)=n]=\frac{(\bar{\eta}A)^n e^{-(\bar{\eta}A)}}{n!}. \tag{2}$$

An operational definition of the average station separation, d, is given by regarding A in terms of an equivalent rectangular region divided into N equal, square cells. It follows that $\bar{\eta}=1/d^2$. Further, the average number of stations within a radius ρ about an arbitrary point in the field is

$$N(\rho)=\int_0^{2\pi}\int_0^{\rho}\bar{\eta}r\,dr\,d\theta=\frac{\pi\rho^2}{d^2}. \tag{3}$$

If attention is restricted to integers $N(\rho_m)=m$, then

$$\rho_m^2=m\,\frac{d^2}{\pi}=m\rho_1^2. \tag{4}$$

As will be shown below, ρ_m^2 is to be interpreted as the expected value of the square of the distance to the mth station, not as the square of the average distance.

If A is a circle described by the radius, r, from an arbitrary point in the field, then eq (2) can be written as

$$P[N(r)=n]=\left(\frac{r^2}{\rho_1^2}\right)^n\frac{e^{-r^2/\rho_1^2}}{n!}. \tag{5}$$

The probability that n or more stations are within r is

$$P[N(r)\geqq n]=P[N(r)\geqq n-1]-P[N(r)=n-1]. \tag{6}$$

Thus,

$$P[N(r)\geqq 1]=1-P[N(r)=0]=1-e^{-r^2/\rho_1^2} \tag{7}$$

and

$$P[N(r)\geqq n]=1-e^{-r^2/\rho_1^2}\sum_{k=0}^{n-1}\frac{1}{k!}\left(\frac{r^2}{\rho_1^2}\right)^k. \tag{8}$$

This is a cumulative distribution in r. The probability density for the distance to the nth station is then given by

$$f_n(r)=\frac{d}{dr}P[N(r)\geqq n]. \tag{9}$$

This can be shown by induction to be

$$f_n(r)=\frac{2r^{2n-1}}{(n-1)!\rho_1^{2n}}\,e^{-r^2/\rho_1^2};\ n=1,2,3,\ldots. \tag{10}$$

[1] Research supported by U.S. Army Electronics Command, Contract No. DAA07-69-C-0062 under Project THEMIS

The distribution is normalized for each n. The average distance to the nth station is

$$\bar{r}_n=\int_0^\infty rf_n(r)dr=\frac{(2n-1)!!}{2^n(n-1)!}d \tag{11}$$

where $(2n-1)!!=1\cdot3\cdot5\cdots(2n-1)$. The variance is

$$\overline{r_n^2}=\int_0^\infty r^2f_n(r)dr=n\rho_1^2, \tag{12}$$

as expected from eq (4).

3. VERIFICATION AND CONCLUSIONS

To verify eq (10), we chose station locations with a pseudo-random number generator in a square domain of dimensions 100 by 100 in arbitrary units. Another point was chosen at random within a centered 20 by 20 subregion. Distances for the closest five stations were collected and stratified in intervals of 0.05 d. For the computations here, $d=4$ units. The experiment was repeated 21,000 times. In each instance, a new station array and reference point were generated. The ensemble of realizations was used in lieu of a larger domain.

The experimental results are summarized in table 1. The average distance and variance for each ordered station are shown as ratios with the values expected from eq (11) and (12). The largest discrepancy is 0.51 percent. The results of a χ^2-test of goodness-of-fit are also shown. Class intervals were grouped on the wings of the distributions so that a total of 31 intervals, or 30 degrees of freedom, were obtained for each station. Since the 5 percent critical value of χ^2 for 30 degrees of freedom is $\chi_c^2=43.77$, there would be no basis for rejecting the hypothesis that eq (10) describes the distance distribution.

TABLE 1.—*Ratios of the first and second moments observed with theoretical values obtained from the probability density. χ^2-values for goodness-of-fit over 31 class intervals are shown for the first five ordered stations.*

Station number	Observed mean / Theoretical mean	Observed variance / Theoretical variance	χ^2
1	1. 0003	1. 0034	23. 4
2	0. 9974	0. 9949	28. 6
3	0. 9979	0. 9956	26. 0
4	0. 9983	0. 9969	22. 0
5	0. 9977	0. 9959	32. 8

ACKNOWLEDGMENT

The author is indebted to R. J. Serfling of the Statistics Department, Florida State University, for suggesting the approach used here.

REFERENCES

Parzen, Emanuel, *Modern Probability Theory and Its Applications*, John Wiley & Sons, New York, N.Y., **1960**, 464 pp.

Stephens, J. J., and Polan, A. L., "Spectral Modification by Objective Analysis," *Monthly Weather Review*, Vol. 99, No. 5, May **1971**, pp. 374–378.

Stephens, J. J., and Stitt, J. M., "Optimum Influence Radii for Interpolation With the Method of Successive Corrections," *Monthly Weather Review*, Vol. 98, No. 9, Sept. **1970**, pp. 680–687.

[*Received April 12, 1971; revised June 24, 1971*]