Mathematical Geology, Vol. 11, No. 6, 1979

Some Exact Sampling Distributions for Variogram Estimators[1]

Bruce M. Davis[2] and Leon E. Borgman[3]

For equally spaced observations from a one-dimensional, stationary, Gaussian random function, the characteristic function of the usual variogram estimator $\hat{\gamma}_k$ for a fixed lag k is derived. Because the characteristic function and the probability density function form a Fourier integral pair, it is possible to tabulate the sampling distribution of a function of a $\hat{\gamma}_k$ using either analytic or numerical methods. An example of one such tabulation is given for an underlying model that is simple transitive. KEY WORDS: geostatistics, variogram estimation, sampling distributions.

INTRODUCTION

The variogram is the function used to estimate and model the intercorrelation structure of a regionalized variable. On the basis of estimates of the variogram made at different lags and orientations (the variogram being a function of a vector argument), a model is chosen for the underlying variogram of the random function assumed to have produced the observed data. A few articles concerning the choice of the model on the basis of observed data have appeared in the literature (e.g., David, 1975; Sabourin, 1975).

The exact sampling distribution for the estimator of the variogram at some lag k may be useful in a model selection procedure or for any of several other reasons.

In general, deriving the sampling distribution of an estimator may be quite difficult; however, if three assumptions are imposed, the problem may be greatly simplified. These assumptions are (1) The randon function is normal, (2) The random function is a covariance stationary, second-order process, (3) The random function exists in one dimension only.

In the sections that follow, the theoretical basis for the derivation of the sampling distribution of a function of the variogram estimator is developed, and the algorithm used to produce tables of the distribution of the statistic is

[1] Manuscript received 4 August 1978; revised 30 April 1979.

[2] Tennessee Valley Authority, Nuclear Raw Materials Branch, P.O. Box 2957, Casper, Wyoming 82602.

[3] Department of Statistics, University of Wyoming, Laramie, Wyoming 82070.

0020-5958/79/1200-0643$03.00/0

briefly described. A set of tables for one such distribution appears in the Appendix.

THE RANDOM FUNCTION

Let $V(x)$ be a normal, second-order, covariance stationary random process in one dimension observed at equally spaced points $x=n\Delta$ for $n=1, 2, \ldots, N+k$. Let $v_n=v(n\Delta)$, $v=(v_1, v_2, \ldots, v_N+k)'$ and let $\mu=(\mu, \mu, \ldots, \mu)'$ where $E[v_n]=\mu$ for all n and μ is a $N+k$ vector. The covariance matrix

$$C=E[(v-\mu)(v-\mu')] \tag{1}$$

$$=E[vv']-\mu\mu' \tag{2}$$

Now define

$$W_n=\frac{v_n-v_{n+k}}{(2N\sigma^2)^{\frac{1}{2}}}, \qquad \text{where } \sigma^2 \text{ is known} \tag{3}$$

If the variogram of $V(x)$ at a lag of $k\Delta$ is denoted γ_k, then consider

$$Y_{N.k}=\sum_{n=1}^{N} W_n^2 \tag{4}$$

$$=\frac{1}{2N\sigma^2}\sum_{n=1}^{N}(v_n-v_{n+k})^2 \tag{5}$$

$$=\hat{\gamma}_k/\sigma^2 \tag{6}$$

where $\hat{\gamma}_k$ is the estimator for the variogram at lag kx, γ_k.

THE CHARACTERISTIC FUNCTION

One way of obtaining the probability law for $\hat{\gamma}_k/\sigma^2$ would be to find its characteristic function and then take the inverse Fourier transform of that function to find the associated probability density function.

To such an end, let A be a matrix of dimensions $n\times(N+k)$ with $A_{ij}=1$ for $j=i$, $A_{ij}=-1$ for $j=i+k$, and $A_{ij}=0$ otherwise. If W is the vector $(W_1, W_2, \ldots, W_n)'$, then it follows that

$$W=[1/(2N\sigma^2)^{\frac{1}{2}}]Av \tag{7}$$

From (3) and the definition of v_n

$$E[W_n]=E[(v_n-v_{n+k})/(2N\sigma^2)^{\frac{1}{2}}]=0 \tag{8}$$

The covariance matrix for W is

$$C_*=E[WW'] \tag{9}$$

$$=(1/2N\sigma^2)E[Avv'A'] \tag{10}$$

$$=(1/2N\sigma^2)AE[vv']A' \tag{11}$$

From (2)

$$C^*=(1/2N\sigma^2)A[C+\mu\mu']A' \tag{12}$$

The characteristic function for $\hat{\gamma}_k/\sigma^2$ is

$$\phi_{\hat{\gamma}_k/\sigma^2}(u)=E\left[\exp\left\{iu\sum_{n=1}^{N}W_n^2\right\}\right] \tag{13}$$

$$=E[\exp\{iuW'W\}] \tag{14}$$

Since the process $V(x)$ is normal, W is multivariate normal with mean vector

$$\mu_W=0 \tag{15}$$

and covariance matrix C^* as given in (12).

Theorem 1. If $V(x)$ is a normal, second-order, covariance stationary random process

$$\phi_{\hat{\gamma}_k/\sigma^2}(u)=(|C^*|)^{\frac{1}{2}}\cdot(|C^{*-1}-2iuI|)$$

Proof. If $V(x)$ satisfies the conditions of the theorem, W is N-variate normal with a mean vector as given in (15) and a covariance matrix as given in (12). Therefore, from (14)

$$\begin{aligned}\phi_{\hat{\gamma}_k/\sigma^2}(u)&=E[\exp\{iuW'W\}]\\&=\int e^{iuW'W}\cdot((2\pi)^N\cdot|C^*|)^{-\frac{1}{2}}e^{-(\frac{1}{2})W'C^{*-1}W)}dW\\&=\int((2\pi)^N\cdot|C^*|)^{-\frac{1}{2}}e^{-\frac{1}{2}(W'[C^{*-1}-2iuI]W)}dW\\&=(|C^*|)^{-\frac{1}{2}}(|C^{*-1}-2iuI|)^{-\frac{1}{2}}\end{aligned}$$ Q.E.D.

THE PROBABILITY DENSITY FUNCTION

It is well known (see for instance, Loeve, 1960, page 185 and 188) that the characteristic function and the probability density function form a Fourier integral pair, that is, if $f(x)$ denotes the density function and $\phi(u)$ denotes the characteristic function of a random variable X

$$\phi(u)=E[e^{iuX}] \tag{16}$$

$$=\int_{-\infty}^{\infty}e^{iux}f(x)\,dx \tag{17}$$

and

$$f(x)=E[e^{-iuX}] \tag{18}$$

$$=(1/2\pi)\int_{-\infty}^{\infty}e^{-iux}\phi(u)\,du \tag{19}$$

Generally, the form of $\phi_{\hat{\gamma}_k/\sigma^2}(u)$ causes (19) to be analytically intractable. No analytic expression of the probability density function for $\hat{\gamma}_k/\sigma^2$ is generally available. However, numerical procedures are available so that the density function may be tabulated.

Using the procedure discussed in Borgman (1977), the characteristic function, $\phi_{\hat{\gamma}_k/\sigma^2}(u)$ may be computed for values $m\Delta u$, $m=0, 1, 2, \ldots, N^*-1$ where N^* is a large number and Δu is an increment on the u-axis. The inverse Fourier transform of the digitized values of $\phi_{\hat{\gamma}_k/\sigma^2}(u)$ may be obtained by use of the Finite Fourier Transform (FFT) algorithm (Rao, 1975) to get a discrete version of the density function

$$f_{\hat{\gamma}_k/\sigma^2}(y) \tag{20}$$

By summation then, a discrete version of the cumulative distribution function

$$F_{\hat{\gamma}_k/\sigma^2}(y) \tag{21}$$

may also be found. From Borgman (1977), if the side condition

$$(\Delta x)(\Delta u)=2\pi/N^*$$

is imposed the numerical analogs to (17) and (19) are

$$\phi_m=\Delta x \sum_{j=0}^{N^*-1} f_j\, e^{i2\pi jm/N^*} \tag{22}$$

$$f_j=\frac{\Delta u}{2\pi}\sum_{m=0}^{N^*-1} \phi_m\, e^{-i2\pi jm/N^*} \tag{23}$$

For the application (23) is the expression of interest. To approximate the probability density $f_{\hat{\gamma}_k/\sigma^2}(y)$, the series in terms of f_j is

$$f_j=\begin{cases} f(0)/2, & \text{if } j=0 \\ f(j\Delta y), & \text{if } j>0 \end{cases} \tag{24}$$

for $0\leqslant i\leqslant N^*-1$. The discrete approximation to the distribution function (21) is

$$F_j=\sum_{i=0}^{j} f_i \Delta y \tag{25}$$

where

$$F_j=P[Y\leq(j+0.5)\Delta y]=F_{\hat{\gamma}/\sigma^2}((j+0.5)\Delta y) \tag{26}$$

A computer program was developed that computes ϕ_m for $0\leqslant m\leqslant N^*-1$ given the simple transitive model

$$\gamma(k)=\begin{cases} a|k| & k<h_0, \\ \sigma^2=1.0 & k\geq h_0, \end{cases} \quad \text{where } a \text{ is a parameter and } h_0 \text{ is the range} \tag{27}$$

for the underlying variogram.

This means that the values v_n are observations from a random function having true variogram as specified by (27).

The program uses subroutine FFT2 (Borgman, 1977) to solve for f_j and F_j. The output of the program is in the F_j tables in the appendix.

A CHECK ON THE TABLES

To check the algorithm which generated the tables of the appendix, a simple example was produced. For this example the values of the parameters were

$$N=2,\ k=1,\ \sigma^2=1.0$$
$$(k\Delta x)=\begin{cases}\frac{1}{2}|k| & |k|<2.0\\ 1 & |k|\geq 2.0\end{cases} \tag{28}$$

where $\Delta=1.0$. With this variogram the matrix C of (3) became

$$C=\begin{pmatrix}1.0 & 0.5 & 0.0\\ 0.5 & 1.0 & 0.5\\ 0.0 & 0.5 & 1.0\end{pmatrix} \tag{29}$$

The matrix A was

$$A=\begin{pmatrix}1 & -1 & 0\\ 0 & 1 & -1\end{pmatrix} \tag{30}$$

Hence, the covariance matrix C^* was

$$C^*=\begin{pmatrix}\frac{1}{4} & 0\\ 0 & \frac{1}{4}\end{pmatrix} \tag{31}$$

From (31), W_1^2 and W_2^2 were independent, and the characteristic function $\phi_{\hat{\gamma}_k/\sigma^2}(u)$ was

$$\phi_{\hat{\gamma}_k/\sigma^2}(u)=(1-\tfrac{1}{2}iu)^{-1} \tag{32}$$

This characteristic function was recognized as that of 1/4 of a χ-square random variable with two degrees of freedom. It was, therefore, possible to check the values obtained by the computer program against 1/4 times the values in the χ-square table with two degrees of freedom. The tabulated values (Snedecor and Cochran, 1967) times 1/4 gave the values listed in Table 1. The values obtained from the program using $N^*=4096$ and $f=0.25$ were listed in Table 2. The values of the two tables agree quite closely.

AREAS FOR FURTHER STUDY

Tabulations of F_j using other models for the underlying variogram may be of

Table 1. $\frac{1}{4}\chi^2_{2,p}$ **Such That** $P(\chi^2_2/4 < \frac{1}{4}x^2_{2,p}) = p$

p	$\frac{1}{4}p$
0.005	0.003
0.01	0.005
0.025	0.014
0.05	0.025
0.10	0.053
0.90	1.153
0.95	1.498
0.975	1.845
0.99	2.303
0.995	2.650

Table 2. $y_{2,1,p}$ **Such That** $P(Y_{2,1} < y_{2,1,p}) = p$

p	$y_{2,1,p}$
0.005	0.003
0.01	0.005
0.025	0.013
0.05	0.025
0.10	0.053
0.90	1.151
0.95	1.499
0.975	1.847
0.99	2.309
0.995	2.661

some use. The process would proceed as developed in the text. Further, it may be possible to derive exact sampling distributions of variogram estimators for two- and three-dimensional processes by the use of a multidimensional FFT algorithm. These would be areas for further research as all possibilities of analytical and numerical intractabilities have not been investigated.

ACKNOWLEDGMENT

The reviewer made several comments that improved the presentation. The authors would like to express their gratitude to this individual and to Cheryl Carroll for assistance in preparing the manuscript.

REFERENCES

Borgman, L. E., 1977, Some new techniques for hurricane risk analysis: 9th Annual OTC Conference Paper, Houston, Texas, May 2–5.

David, M., 1975, Geostatistical ore reserve estimation (Review copy): Ecole Polytechnique de Montreal, Quebec, Canada, 550 p.

Loeve, M., 1960, Probability theory (2nd ed.): D. Van Nostrand Co. Inc., New York, New York, 685 p.

Rao, K. R., 1975, Orthogonal transformation for digital signal processing: Springer-Verlag, Berlin, 263 p.

Sabourin, R., 1975, Application of two methods for the interpretation of the underlying variogram, *in* Advanced geostatistics in the mineral industry, Eds.: M. Guarascio, M. David, and C. Huijbregts: N.A.T.O. Advanced Study Institutes Series, D. Reidell, Boston, p. 101–112.

Snedecor, G. and W. G. Cochran, 1967, Statistical methods: Iowa State University Press, Ames, Iowa, 593 p.

APPENDIX

Explanation of the Tables

The tables of this appendix are a tabulation of the sampling distribution of the statistic

$$Y_{N,k} = \sum_{n=1}^{N} W_n^2$$

derived in the paper. The underlying model assumed is a simple transitive model with range h_0 and sill $\sigma^2 = 1.0$, that is

$$\gamma(k) = \begin{cases} a|k| & |k| < h_0 \\ \sigma^2 = 1.0 & |k| \geq h_0 \end{cases}$$

The tables give $Y_{n,k,h_0,p}$ where $P(Y_{n,k,} < y_{n,k,h_0 p}) = p$ for $N = 2,5(5)\ 25$, $k = 1(1)\{h_0 - 1\}, h_0 = 2.0$ and other values less than 10.0, and $p = 0.01,\ 0.025,\ 0.05,\ 0.10,\ 0.90,\ 0.95,\ 0.975,\ 0.99$.

The values of N and h_0 are listed at the head of each table. The values of p are listed horizontally under the values of N and h_0, and the values of k are listed down the left-hand side of the table.

An Example of Table Use

To find $y_{10,3,5,0.95}$ using the tables, enter the table headed "Zone of Influence = 5, $N = 10$," go down the column labeled K to the value 3. Read across the row $K = 3$ and under the column $p = 0.95$ to find $y_{10,3,5,0.95} = 3.952$.

Sampling distribution of $Y_{Nk} = \sum_{n=1}^{N} W_n^2$

The tabulated value $= Y(N,K,P)$, the zone of influence $= 2.0$, $N = 2$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.005	0.013	0.025	0.053	1.151	1.499	1.847	2.309

The tabulated value $= Y(N,K,P)$, the zone of influence $= 2.0$, $N = 5$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.008	0.020	0.043	0.119	3.915	3.960	3.980	3.992

The tabulated value $= Y(N,K,P)$, the zone of influence $= 3.0$, $N = 5$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.005	0.014	0.028	0.074	3.943	3.973	3.986	3.995
2	0.010	0.026	0.064	0.223	3.888	3.948	3.975	3.990

The tabulated value $= Y(N,K,P)$, the zone of influence $= 4.0$, $N = 5$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.004	0.010	0.022	0.053	3.957	3.979	3.990	3.996
2	0.008	0.020	0.045	0.200	3.918	3.961	3.980	3.992
3	0.011	0.029	0.079	0.277	3.880	3.994	3.974	3.989

The tabulated value $= Y(N,K,P)$, the zone of influence $= 2.0$, $N = 10$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.008	0.020	0.039	0.080	3.914	3.960	3.980	3.992

The tabulated value = $Y(N,K,P)$, the zone of influence = 3.0, $N = 10$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.005	0.013	0.025	0.054	3.945	3.974	3.987	3.995
2	0.011	0.026	0.053	0.114	3.890	3.947	3.975	3.990

The tabulated value = $Y(N,K,P)$, the zone of influence = 4.0, $N = 10$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.004	0.010	0.020	0.040	3.959	3.980	3.990	3.996
2	0.008	0.020	0.039	0.081	3.916	3.961	3.980	3.992
3	0.012	0.028	0.059	0.136	3.876	3.941	3.972	3.989

The tabulated value = $Y(N,K,P)$, the zone of influence = 5.0, $N = 10$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.003	0.008	0.016	0.031	3.968	3.984	3.992	3.997
2	0.006	0.016	0.032	0.067	3.935	3.969	3.984	3.994
3	0.009	0.023	0.047	0.100	3.983	3.952	3.977	3.991
4	0.012	0.030	0.063	0.165	3.870	3.938	3.971	3.988

The tabulated value = $Y(N,K,P)$, the zone of influence = 2.0, $N = 15$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.007	0.018	0.036	0.072	3.921	3.963	3.982	3.993

The tabulated value = $Y(N,K,P)$, the zone of influence = 3.0, $N = 15$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.005	0.012	0.023	0.048	3.949	3.976	3.988	3.995
2	0.010	0.024	0.051	0.106	3.898	3.951	3.976	3.990

The tabulated value = $Y(N,K,P)$, the zone of influence = 5.0, $N = 15$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.003	0.007	0.014	0.028	3.971	3.986	3.993	3.997
2	0.006	0.015	0.030	0.063	3.939	3.971	3.985	3.994
3	0.009	0.023	0.048	0.104	3.905	3.954	3.978	3.991

The tabulated value = $Y(N,K,P)$, the zone of influence = 8.0, $N = 15$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.003	0.004	0.009	0.018	3.982	3.991	3.996	3.998
2	0.004	0.009	0.019	0.038	3.961	3.981	3.991	3.996
3	0.006	0.015	0.029	0.063	3.940	3.972	3.986	3.994
4	0.007	0.019	0.037	0.080	3.920	3.963	3.981	3.993
5	0.009	0.023	0.048	0.103	3.888	3.951	3.977	3.991
6	0.011	0.028	0.059	0.146	3.879	3.944	3.973	3.989
7	0.013	0.032	0.066	0.154	3.858	3.936	3.969	3.987

The tabulated value = $Y(N,K,P)$, the zone of influence = 3.0, $N = 20$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.007	0.017	0.033	0.066	3.927	3.996	3.983	3.993

The tabulated value = $Y(N,K,P)$, the zone of influence = 3.0, $N = 20$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.004	0.011	0.022	0.044	3.954	3.979	3.989	3.996
2	0.009	0.023	0.047	0.099	3.904	3.954	3.977	3.991

The tabulated value = $Y(N,K,P)$, the zone of influence = 5.0, $N = 20$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.003	0.007	0.013	0.026	3.974	3.987	3.994	3.997
2	0.006	0.014	0.028	0.058	3.943	3.973	3.986	3.994
3	0.009	0.023	0.045	0.097	3.909	3.956	3.979	3.991
4	0.012	0.029	0.060	0.124	3.872	3.940	3.971	3.988

The tabulated value = $Y(N,K,P)$, the zone of influence = 8.0, $N = 20$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.003	0.004	0.008	0.016	3.983	3.992	3.996	3.998
2	0.003	0.009	0.018	0.035	3.964	3.982	3.991	3.997
3	0.006	0.014	0.027	0.059	3.994	3.973	3.986	3.995
4	0.007	0.018	0.036	0.077	3.921	3.963	3.982	3.993
5	0.010	0.023	0.049	0.110	3.901	3.953	3.977	3.991
6	0.011	0.027	0.057	0.125	3.880	3.944	3.973	3.989
7	0.013	0.032	0.065	0.140	3.855	3.934	3.968	3.987

The tabulated value = $Y(N,K,P)$, the zone of influence = 10.0, $N = 20$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.002	0.003	0.006	0.013	3.987	3.994	3.997	3.999
2	0.003	0.007	0.014	0.027	3.972	3.986	3.993	3.997
3	0.004	0.011	0.022	0.045	3.955	3.979	3.989	3.996
4	0.006	0.015	0.030	0.064	3.938	3.971	3.985	3.994
5	0.007	0.018	0.037	0.079	3.921	3.963	3.982	3.993
6	0.009	0.023	0.045	0.097	3.895	3.953	3.978	3.991
7	0.010	0.026	0.055	0.128	3.888	3.948	3.975	3.990
8	0.012	0.029	0.061	0.140	3.871	3.940	3.972	3.988
9	0.013	0.032	0.067	0.148	3.854	3.934	3.968	3.987

The tabulated value = $Y(N,K,P)$, the zone of influence = 2.0, $N = 25$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.006	0.016	0.031	0.063	3.936	3.969	3.985	3.994

The tabulated value = $Y(N,K,P)$, the zone of influence = 3.0, $N = 25$

K \ P	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.004	0.010	0.020	0.040	3.957	3.979	3.990	3.996
2	0.009	0.023	0.045	0.093	3.911	3.956	3.979	3.991