ASYMPTOTIC EFFICIENCY IN POLYNOMIAL ESTIMATION[1]

By Paul G. Hoel

University of California, Los Angeles

1. Summary. Asymptotic formulas are obtained for the generalized variance of the least squares estimates in polynomial regression under the assumption that the basic random variables are those of a stationary stochastic process, or a slight generalization of such a process. These formulas are used to study the information obtained by increasing the number of observational points in an interval and by increasing the length of the interval.

2. Introduction. In an earlier paper [1], some limited results were obtained on the increased efficiency of estimation in polynomial regression due to increasing the number of observational points, under the assumption that the basic random variables are correlated. These results were for two special stochastic processes only. In this paper, somewhat more general stochastic processes are studied and corresponding asymptotic formulas are obtained.

The same notation will be used here as in [1]. Thus, $y_1, y_2, \cdots, y_n$ will denote random variables associated with the fixed values $x_1, x_2, \cdots, x_n$, and the regression polynomial will be denoted by

$$E(y_i) = \beta_0 + \beta_1 x_i + \cdots + \beta_k x_i^k .$$

For convenience the interval $(0, a)$ will be chosen as the interval over which observations are to be taken. Furthermore, in the development of the theory, the observation points $x_1, x_2, \cdots, x_n$ will be chosen to be the n equally spaced points given by the formula $x_i = i\delta$, where $\delta = a/n$.

The variables $y_1, y_2, \cdots, y_n$ will be assumed to be those of a stationary stochastic process. Thus, the y's possess a common variance, and the correlation between y_i and y_j is a function of $|i - j|\delta$ only. As a result, the covariance matrix S can be written in the form

$$S = \sigma^2 \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{n-1} \\ r_1 & 1 & r_1 & \cdots & r_{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{n-1} & r_{n-2} & r_{n-3} & \cdots & 1 \end{bmatrix}.$$

Here r_j denotes the correlation coefficient for two variables whose x values are $j\delta$ units apart.

As before, it is necessary to introduce the spacing matrix X given by the formula

Received October 22, 1960; revised March 10, 1961.

[1] This research was supported by means of an Air Force Research Grant.

$$X = \begin{bmatrix} 1 & \delta & \delta^2 & \cdots & \delta^k \\ 1 & 2\delta & (2\delta)^2 & \cdots & (2\delta)^k \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & n\delta & (n\delta)^2 & \cdots & (n\delta)^k \end{bmatrix}.$$

The measure of efficiency in the estimation of the β's that will be used here is the generalized variance, or, equivalently, the square of the volume of the ellipsoid of concentration. The generalized variance for best unbiased linear estimates is expressible by means of a well known formula [1] in terms of the matrices X and S.

3. Least squares estimates. From a theorem in the book of Grenander and Rosenblatt [2], it follows that the least squares estimates of the coefficients in polynomial regression are asymptotically efficient for stationary processes. Therefore, in studying asymptotic efficiency, least squares estimates may be used in place of Markoff estimates, provided one is dealing with stationary processes. Since least squares estimates possess a simpler generalized variance formula than Markoff estimates, it is convenient to work with them in studying the asymptotic efficiency of various spacing designs. For least squares estimates, the generalized variance of polynomial regression coefficients is given [2] by the formula

$$\text{G.V.} = \frac{|X'SX|}{|X'X|^2}. \tag{1}$$

Now consider any continuous correlation function $\rho(t)$ defined over the closed interval $[0, a]$. Since it may be approximated arbitrarily closely by a finite series of the form

$$\rho(t) = \sum_{m=1}^{N} c_m \exp\{-\alpha_m t\}, \tag{2}$$

where $\alpha_m > 0$, it will be assumed that the correlation function is of this type. Because $\rho(0) = 1$, it is necessary that $\sum c_m = 1$. In terms of this correlation function the value of r_j will be given by

$$r_j = \rho(j\delta) = \sum_{m=1}^{N} c_m \exp\{-\alpha_m j\delta\}.$$

Then S assumes the form $S = \sigma^2(w_{ij})$ where

$$w_{ij} = \sum_{m=1}^{N} c_m \exp\{-\alpha_m \delta|i - j|\}.$$

Let $S_m = (w_{ij}^{(m)})$ where $w_{ij}^{(m)} = \exp\{-\alpha_m\delta|i - j|\}$. Then S may be written as

$$S = \sigma^2 \sum_{m=1}^{N} c_m S_m.$$

As a result

$$X'SX = \sigma^2 \sum_{m=1}^{N} c_m X'S_m X. \tag{3}$$

Consider the typical term, $a_{i+1,j+1}^{(m)}$, in $X'S_mX$. Postmultiplying S_m by X and premultiplying the result by X' will show that $a_{i+1,j+1}^{(m)}$ can be expressed in the form

$$a_{i+1,j+1}^{(m)} = \sum_{x=1}^{n} \sum_{y=1}^{n} (x\delta)^i (y\delta)^j e^{-\alpha_m |x\delta - y\delta|}.$$

If the substitutions $u = x\delta$ and $v = y\delta$ are made, it will be seen that this sum possesses the same asymptotic value with respect to $1/\delta$ as the integral

$$\left(\frac{1}{\delta}\right)^2 \int_0^a \int_0^a u^i v^j e^{-\alpha_m |u-v|} \, du \, dv. \tag{4}$$

The typical term in $X'SX$ will be denoted by $b_{i+1,j+1}$. From (3) it follows that

$$b_{i+1,j+1} = \sigma^2 \sum_{m=1}^{N} c_m a_{i+1,j+1}^{(m)}, \tag{5}$$

and therefore that the asymptotic value of $b_{i+1,j+1}$ is obtained by substituting the asymptotic value of $a_{i-1,j+1}^{(m)}$ into (5). In view of formulas (2) and (4), this asymptotic value may be written in the form

$$b_{i+1,j+1} \sim \left(\frac{1}{\delta}\right)^2 \sigma^2 \int_0^a \int_0^a u^i v^j \rho(u - v) \, du \, dv. \tag{6}$$

Similar considerations will show that the typical term in $X'X$ is given by the single sum $\sum_{x=1}^{n} (x\delta)^i (x\delta)^j$, which possesses the asymptotic value

$$\left(\frac{1}{\delta}\right) \int_0^a u^{i+j} \, du = \frac{a^{i+j+1}}{\delta(i + j + 1)}. \tag{7}$$

It now follows from (1), (6), and (7) that the asymptotic value of the generalized variance is given by

$$\text{G. V.} \sim \frac{\left| \left(\frac{\sigma}{\delta}\right)^2 \int_0^a \int_0^a u^i v^j \rho(u - v) \, du \, dv \right|}{\left| \frac{1}{\delta} \frac{a^{i+j+1}}{i + j + 1} \right|^2}.$$

Since the factors in $1/\delta$ cancel, this asymptotic expression reduces to

$$\text{G. V.} \sim \frac{\left| \sigma^2 \int_0^a \int_0^a u^i v^j \rho(u - v) \, du \, dv \right|}{\left| \frac{a^{i+j+1}}{i + j + 1} \right|^2}. \tag{8}$$

4. Nonconstant variance. The preceding results were based on the assumption that the process is stationary. This assumption is certainly a realistic one for many applications, at least as far as the correlation function is concerned. A more general situation, in which the variance is assumed to be a continuous function of time in the closed interval $[0, a]$, can be treated by the same methods as those just employed and will be found to yield similar conclusions with respect to asymptotic efficiency.

The demonstration of this last statement can be carried out by considering σ_t as a polynomial in t. The resulting change in the matrix S_m will change the integral (4) to the integral

$$\left(\frac{1}{\delta}\right)^2 \int_0^a \int_0^a u^i v^j e^{-\alpha_m|u-v|} \sigma_u \, \sigma_v \, du \, dv, \tag{9}$$

where, say,

$$\sigma_u = \gamma_0 + \gamma_1 u + \cdots + \gamma_s u^s.$$

The substitution of this quantity in (9) will show that (9) reduces to

$$\left(\frac{1}{\delta}\right)^2 \sum_{p=0}^{s} \sum_{q=0}^{s} \gamma_p \gamma_q \int_0^a \int_0^a u^{i+p} v^{j+q} e^{-\alpha_m|u-v|} \, du \, dv. \tag{10}$$

As a result, the asymptotic value of (10) is of the same form as for (4), and therefore one would expect the same type of asymptotic efficiency properties to hold. Such properties will be considered in the next section.

5. Adding observations. The result given by (8) shows that when a large number of observations has been made in an interval the amount of information gained by taking additional observations is negligible. Thus, if the number of equally spaced points in an interval is doubled, which means that δ must be replaced by $\delta/2$, the same asymptotic value of the generalized variance is obtained because (8) does not depend upon δ. This holds regardless of the nature of the correlation function, provided that it is continuous. It holds not only for any stationary process but also, in view of formula (10), for a stationary process that has been modified to permit the variance to be any continuous function of t.

If the y's are independent random variables, the generalized variance will approach zero as δ approaches zero, and therefore it certainly pays to add observations in this case. In view of this fact, it is clear that the size of the sample needed before one can conclude that it is hardly worth while taking additional observations depends rather heavily upon the nature of the correlation function. For the purpose of observing how the value of the generalized variance changes as the nature of the correlation function changes, consider the special correlation function $\rho(t) = e^{-\alpha t}$ that was considered in [1], and assume that $\sigma = 1$. Suppose the value of α is changed to the value 2α. This is equivalent to squaring the value of the correlation coefficient between any pairs of y values, and hence in weakening the correlation relationship to this degree. For any particular value of a and α this effect of doubling α can be determined numerically by means of formula (8); however, it is difficult to make such a comparison for a general α unless α is very large. Therefore, consider next an approximation to (8) which is valid for large values of α.

Let

$$I_{ij} = \int_0^a \int_0^u u^i v^j e^{-\alpha(u-v)}\, dv\, du.$$

The value of $b_{i+1,j+1}$ in (6) will then be given by the quantity $(I_{ij} + I_{ji})/\delta^2$. Repeated integration by parts in the first integration will show that I_{ij} can be expressed in the form

$$I_{ij} = \frac{1}{\alpha}\left\{\frac{a^{i+j+1}}{i+j+1} - \frac{j}{\alpha}\frac{a^{i+j}}{i+j} + \frac{j(j-1)}{\alpha^2}\frac{a^{i+j-1}}{i+j-1} - \cdots (-1)^j \frac{j(j-1)\cdots 1}{\alpha^j}\frac{a^{i+j}}{i+1}\right\} + (-1)^{j+1}\frac{j(j-1)\cdots 1}{\alpha^{j+1}}\int_0^a u^i e^{-\alpha u}\, du.$$

For large α the first term will dominate this expression; therefore I_{ij} may be approximated by

$$I_{ij} \approx \frac{1}{\alpha}\frac{a^{i+j+1}}{i+j+1}.$$

The double integral in (8) will therefore be approximated by twice this value; consequently, for large α, the generalized variance in (8) may be approximated by

$$\text{G. V.} \approx \frac{\left|\frac{2}{\alpha}\frac{a^{i+j+1}}{i+j+1}\right|}{\left|\frac{a^{i+j+1}}{i+j+1}\right|^2} = \frac{(2/\alpha)^{k+1}}{Aa^{(k+1)^2}}, \tag{11}$$

where $A = |1/(i+j+1)|$.

In making comparisons by means of the generalized variance, it is convenient to consider the quantity introduced in [1], namely,

$$\left[\frac{\text{G. V. }(\alpha, a)}{\text{G. V. }(2\alpha, a)}\right]^{1/(k+1)}$$

The value of this quantity gives the number of replications of an experiment in the given interval needed to yield the same estimation information, as measured by the generalized variance, as that obtained through doubling the value of α. From (11) it follows that the value of this quantity is 2 here; therefore when a large number of observations has been made, doubling the value of a large α yields the same estimation information as repeating the experiment. If the typical element in the covariance matrix $(X'X)^{-1}(X'SX)(X'X)^{-1}$ is computed, using the same approximation as before, it will be seen that doubling the value of α multiplies all elements of this matrix by $\frac{1}{2}$; therefore in the sense that the variance of each estimate is multiplied by $\frac{1}{2}$, the efficiency of estimation is doubled through doubling α.

The preceding results show that doubling the number of observation points in an interval gives rise to two counteracting effects. The favorable effect arises from doubling the size of the experiment. The unfavorable effect arises from increasing the correlation between neighboring variables. For large samples, these two effects approximately nullify each other. Since the preceding result, that adding points does not help much here, holds for a large value of α, and hence for a weak correlation relationship, the advantage of adding observations in a fixed interval would be expected to be even less when there exists a strong correlation relationship. Some numerical results in this connection may be found in [1].

6. Extending the interval. A second form of comparison which is of interest in regression problems is that arising when the interval over which observations are to be taken is extended. This comparison for the same correlation function as in Section 5 can be made by replacing a by $2a$ in (8) and then calculating the quantity

$$\left[\frac{\text{G. V. }(\alpha, a)}{\text{G V. }(\alpha, 2a)}\right]^{1/(k+1)}. \tag{12}$$

When α is large, the approximation given in (11) may be used, in which case the value of this quantity will reduce to 2^{k+1}. Thus, when α is large and a large number of observations has been made, doubling the number of equally spaced observations by doubling the length of the interval gives approximately as much estimation information as 2^{k+1} replications of the experiment in the original interval. This result was obtained in [1] by using other methods.

When α is not sufficiently large to justify the use of approximation (11), numerical methods are needed to observe what effect doubling the length of the interval has on the generalized variance. Thus, calculations for $\alpha = 1$, $a = 1$, and $k = 2$ by means of formula (8) yielded the value 15 for the quantity given in (12). Under independence the value would have been 8; therefore there appears to be even greater advantage in extending the interval when strong correlation exists than under independence.

REFERENCES

[1] Paul G. Hoel, "Efficiency problems in polynomial estimation," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1134–1145.

[2] Ulf Grenander and Murray Rosenblatt, *Statistical Analysis of Stationary Time Series*, John Wiley and Sons, New York, 1956.