SPACING OF INFORMATION IN POLYNOMIAL REGRESSION[1]

By A. de la Garza

Carbide and Carbon Chemicals Company

Oak Ridge, Tennessee

1. Summary and Introduction. The purpose of this paper is to investigate a problem in the spacing of information in certain applications of polynomial regression. It is shown that for a polynomial of degree m, the variance-covariance matrix of the estimated polynomial coefficients given by a spacing of information at more than $m + 1$ values of the sure variate can always be attained by spacing the same information at only $m + 1$ values of the sure variate, these spacing values being bounded by the first spacing values. The presented results are of use in experimental design involving polynomial regression when a choice of sure variate values is possible but restricted to a specified range.

Let the polynomial under consideration be

$$P(x) = \alpha_1 + \alpha_2 x + \cdots + \alpha_{m+1} x^m,\ m \geqq 1, \tag{1.1}$$

and let

$$P(x_\epsilon) = y(x_\epsilon) + \delta_\epsilon,\ \epsilon = 1(1)N,\ N \geqq (m + 1).$$

The $y(x_\epsilon)$ are observed uncorrelated variates with random error δ_ϵ having mean zero and finite variance $\sigma_\epsilon^2 > 0$. The x_ϵ are observed variates without error, there being at least $(m + 1)$ distinct x_ϵ.

The following notation is introduced. Let $\bar{x} = (1, x, x^2, \cdots, x^m)$, $X = (\bar{x}_\epsilon)$, $\epsilon = 1(1)N$, and let W be the $N \times N$ diagonal matrix with entry $w_\epsilon = 1/\sigma_\epsilon^2$ in the (ϵ, ϵ) position. w_ϵ will henceforth be referred to as the "information" of $y(x_\epsilon)$, and $Q = \sum w_\epsilon$, $\epsilon = 1(1)N$, will be referred to as the "total information." The matrix $X'WX$ will be called the "information matrix."

The problem is to show that given a spacing of total information Q at locations x_ϵ, $\epsilon = 1(1)N$, $N \geqq (m + 1)$, there being at least $(m + 1)$ distinct x_ϵ, it is always possible to re-space Q at $(m + 1)$ distinct locations r_j, $j = 1(1)(m + 1)$, in such a manner that $\min x_\epsilon \leqq r_j \leqq \max x_\epsilon$, $\epsilon = 1(1)N$, $j = 1(1)(m + 1)$, and $X'WX = R'UR$, with $R'UR$ being the information matrix of the re-spacing. The problem is solved by prescribing a method for finding the required U and R which determine the spacing of the total information.

The motivation for the problem is as follows. In experimentation in the chemical engineering industry, we most often have control over our sure variates. The sure variate x could be the pressure level of our process equipment, and we would be permitted to choose any operating pressure x in the pressure range min x to max x, tolerated by our equipment. Quite often, and in particular with isotopic measurements, laboratory analytical determinations are required for our y-

Received 1/7/53, revised 8/12/53.

[1] Work performed under AEC Contract No. W-7405-eng-26.

variates with the laboratory being the major source of error. With each laboratory determination having variance σ^2, we can request n_x determinations on the material sample taken at sure variate x. Using the average of the laboratory determinations, the corresponding y variate has variance σ^2/n_x. Specifying Q then amounts to specifying total laboratory effort expended on the experiment. It might be set by such usual factors as the dollar allowance on the experiment; if the material is highly radioactive, it might be set by such unusual factors as exposure time allowed the laboratory analysts. Furthermore, in experimentation with fairly large equipment, it is important to minimize the distinct levels of operation, that is, the distinct number of sure x's. The time required to make the change and to reach sufficient equilibrium representing steady-state operation of the process is often long. In any case we lose time, and with production line equipment, we also lose production. These are the reasons for minimizing the distinct number of sure x's in the experiment. The equivalence $X'WX = R'UR$ gives the required minimization. If the functional relationship between y and x is adequately represented by a polynomial of degree m, the equivalence assures that only $(m + 1)$ distinct sure x's are required to maintain the same efficiency of statistical evaluation of the experimental results, since most statistical evaluation will require $(X'WX)^{-1}$, which can now be replaced by $(R'UR)^{-1}$.

It may be seen that such experiments, common in physico-chemical industry, present a formulation and require a mathematical model not found in ordinary regression theory, where usually it is not possible to assign various values to the corresponding y variances.

With the indicated background in mind, the results of this paper find application in experimental design. The determination of a spacing which optimizes some criteria involving the information matrix is made simpler. A familiar example arising in point estimation is minimizing $\vec{p}(X'WX)^{-1}\vec{p}'$ for a specified row vector $\vec{p}$. An example from interpolation is minimizing the maximum of $\vec{\xi}(X'WX)^{-1}\vec{\xi}'$ with $\vec{\xi} = (1, \xi, \xi^2 \cdots \xi^m)$ and $\min x_\epsilon \leqq \xi \leqq \max x_\epsilon$; the extrapolation problem is similar.

The advantage of applying the above result to such problems is that the spacing of information is at once reduced to $(m + 1)$ distinct locations, any larger number being unnecessary. The matrix X then is the matrix of a Vandermonde determinant. The properties of these matrices are well known and attractive. These uses will be illustrated by an example given in Section 4.

2. Some useful relations. Prior to investigating the problem as outlined above, several relations needed later will be developed. First, a convention in subscripts: subsequently, small italic letters will run from 1 to $(m + 1)$, and small Greek letters will run from 1 to N. Capital italic letters will run as indicated.

With the notation of Section 1 and with $\vec{\alpha} = (\alpha_1\ \alpha_2 \cdots \alpha_{m+1})'$, the polynomial (1.1) under consideration is

$$P(x) = \vec{x}\vec{\alpha}. \tag{2.1}$$

Choose $(m + 1)$ distinct numbers z_j. From Lagrange interpolation it follows that

$$P(x) = \frac{(x - z_2)(x - z_3)\cdots(x - z_{m+1})}{(z_1 - z_2)(z_1 - z_3)\cdots(z_1 - z_{m+1})} P(z_1) + \frac{(x - z_1)(x - z_3)\cdots(x - z_{m+1})}{(z_2 - z_1)(z_2 - z_3)\cdots(z_2 - z_{m+1})} \cdot P(z_2) + \cdots + \frac{(x - z_1)(x - z_2)\cdots(x - z_m)}{(z_{m+1} - z_1)(z_{m+1} - z_2)\cdots(z_{m+1} - z_m)} P(z_{m+1}).$$

With an obvious notation,

$$P(x) = \textstyle\sum_j F(x, z_j)P(z_j). \tag{2.2}$$

Consider now that from (2.1) $\vec{P}(z) = Z\vec{\alpha}$, where $\vec{P}(z) = (P(z_1)\ P(z_2) \cdots P(z_{m+1}))'$, $\vec{z} = (1\ z \cdots z^m)$, and $Z = (\vec{z}_j)$. The matrix Z is nonsingular, since its determinant is a Vandermonde determinant not equal to zero due to the z_j being distinct. Thus, $Z^{-1}\vec{P}(z) = \vec{\alpha}$, and for any x,

$$\vec{x}Z^{-1}\vec{P}(z) = \vec{x}\vec{\alpha}. \tag{2.3}$$

Since from (2.2)

$$P(x) = F(x, z)\vec{P}(z), \tag{2.4}$$

with $\vec{F}(x, z) = (F(x, z_1)\ F(x, z_2) \cdots F(x, z_{m+1}))$, it follows from (2.1), (2.3), and (2.4) that

$$\vec{x}Z^{-1}\vec{P}(z) = \vec{F}(x, z)\vec{P}(z). \tag{2.5}$$

Equality of (2.5) for any x implies

$$\vec{x}Z^{-1} = \vec{F}(x, z). \tag{2.6}$$

3. Investigation of the problem. The problem as stated in Section 1 is now investigated. From the results of Section 2, it may be shown that without loss in generality the range of the variable x may be limited to $\min x_\epsilon = -1$ and $\max x_\epsilon = 1$. Another simplification follows. Suppose that some of the x_ϵ are not distinct. Say that $x_1 = x_2 = \cdots = x_K$ with corresponding information w_1, $w_2 \cdots, w_K$. It may be verified directly that the information matrix for w_1, $w_2 \cdots, w_K, w_{K+1}, \cdots w_N$ at $x_1, x_2 \cdots x_K, x_{K+1} \cdots x_N$ is the same as the information matrix for $(w_1 + w_2 + \cdots + w_K), w_{K+1}, \cdots w_N$ at $x_K, x_{K+1}, \cdots x_N$. Such a grouping can be made for all x_ϵ not distinct, thereby reducing the problem to considering only distinct x_ϵ. Finally, for $N = (m + 1)$, there is no problem since the information already is at $(m + 1)$ locations.

The problem may now be re-stated as follows. It must be shown that given a spacing of total information Q at N distinct locations x_ϵ, $\epsilon = 1(1)N$, $N > (m + 1)$, with $\min x_\epsilon = -1$, and $\max x_\epsilon = 1$, it is always possible to re-space Q at $(m + 1)$ distinct locations r_j, $j = 1(1)(m + 1)$, in such a manner that $-1 \leqq r_j \leqq 1$, and $X'WX = R'UR$.

Suppose that R exists. Then,

$$(XR^{-1})'W(XR^{-1}) = U. \tag{3.1}$$

Reference to (2.6) shows that the off-diagonal elements of $(XR^{-1})'W(XR^{-1})$ are proportional to

$$c_{gh} = \sum_\epsilon w_\epsilon \phi_\epsilon \prod_p (x_\epsilon - r_p), \qquad g \neq h, p \neq g, p \neq h, \tag{3.2}$$

with $\phi_\epsilon = \prod_j (x_\epsilon - r_j)$. Since in (3.1), U is diagonal, it is required that c_{gh} be zero for $g \neq h$. This requirement is satisfied if the r_j are determined such that

$$\sum w_\epsilon \phi_\epsilon = 0, \qquad \sum w_\epsilon \phi_\epsilon x_\epsilon = 0, \cdots, \sum w_\epsilon \phi_\epsilon x_\epsilon^{m-1} = 0. \tag{3.3}$$

For reasons that will be discussed later, further constrain the r_j by

$$\sum w_\epsilon \phi_\epsilon x_\epsilon^m = 0. \tag{3.4}$$

By direct expansion,

$$\phi_\epsilon = \beta_1 + \beta_2 x_\epsilon + \cdots + \beta_{m+1} x_\epsilon^m + x_\epsilon^{m+1}. \tag{3.5}$$

Hence, the r_j are the $(m + 1)$ roots of the polynomial

$$\mathcal{P}(r) = \beta_1 + \beta_2 r + \cdots + \beta_{m+1} r^m + r^{m+1}. \tag{3.6}$$

Substituting (3.5) in (3.3) and (3.4), there results

$$\begin{pmatrix} f_0 & f_1 & \cdots f_{m-1} & f_m \\ f_1 & f_2 & \cdots f_m & f_{m+1} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ f_{m-1} & f_m & \cdots f_{2m-2} & f_{2m-1} \\ f_m & f_{m+1} & \cdots f_{2m-1} & f_{2m} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_m \\ \beta_{m+1} \end{pmatrix} + \begin{pmatrix} f_{m+1} \\ f_{m+2} \\ \cdot \\ \cdot \\ \cdot \\ f_{2m} \\ f_{2m+1} \end{pmatrix} = 0, \tag{3.7}$$

with

$$f_L = \sum_\epsilon w_\epsilon x_\epsilon^L, \qquad L = 0(1)(2m + 1).$$

(3.7) is a linear system of $(m + 1)$ equations in $(m + 1)$ unknowns. The square matrix is $X'WX$, which is nonsingular, and hence (3.7) has a unique solution, β_j. This solution is not trivial, since it is readily seen that some f_L, $(m + 1) \leqq L \leqq (2m + 1)$, is not zero. The corresponding r_j are then given by (3.6).

Thus, a method of determining the r_j that satisfy (3.3) and (3.4) has been prescribed. Accordingly, these r_j make c_{gh} zero in (3.2) as required. It will now be shown that they are real and distinct.

That the r_j are real is shown as follows. Since they are the roots of the polynomial (3.6), complex r_j, if any, must occur in conjugate pairs. Say that $r_1 = b_1 + b_2 i$ and $r_2 = b_1 - b_2 i$, with $b_2 \neq 0$, the nature of the remaining roots being unspecified. Since in (3.2), c_{12} is then zero, it follows that

$$\sum_\epsilon w_\epsilon[(x_\epsilon - b_1)^2 + b_2^2](x_\epsilon - r_3)^2 \cdots (x_\epsilon - r_m)^2(x_\epsilon - r_{m+1})^2 = 0. \tag{3.8}$$

Note that all factors in each term of (3.8) are nonnegative, and therefore, equality to zero implies that all terms must be zero. But $[(x_\epsilon - b_1)^2 + b_2^2]$ with $b_2 \neq 0$ never vanishes, and $(x_\epsilon - r_3)^2 \cdots (x_\epsilon - r_m)^2(x_\epsilon - r_{m+1})^2$ can vanish for at most $(m - 1)$ distinct x_ϵ. Since there are at least $(m + 2)$ distinct x_ϵ, it follows that (3.8) cannot be zero, and hence r_1 and r_2 are not complex. The argument is the same for any other pair of roots, and hence, all the r_j are real. That they are distinct is shown similarly. Suppose $r_1 = r_2 = b_1$, and again form c_{12}, which is now given by (3.8) with $b_2 = 0$. The arguments are as before. The terms can now vanish for at most m distinct x_ϵ, but since there are at least $(m + 2)$ distinct x_ϵ, (3.8) with $b_2 = 0$ cannot be zero. Hence, all the r_j are distinct.

Since the r_j are distinct, it follows that the matrix R is nonsingular, and that c_{gh} being zero in (3.2) implies that $(XR^{-1})'W(XR^{-1}) = U$ is a diagonal matrix. Since both $X'WX$ and R are nonsingular, it follows that no diagonal element of U is zero. Further reference to (2.6) shows that the diagonal elements of U are

$$u_h = \sum_\epsilon w_\epsilon \prod_{\substack{j \\ j\neq h}} \frac{(x_\epsilon - r_j)^2}{(r_h - r_j)^2}, \tag{3.9}$$

and hence, all u_h are positive.

Thus, with U given by (3.9), $X'WX = R'UR$. Since the (1,1) element of $X'WX$ is the total information $Q = \sum w_\epsilon$, and since the (1,1) element of $R'UR$ is $\sum u_j$, it follows that $Q = \sum u_j$. Inasmuch as it has been shown that $u_j > 0$, the u_j may be considered a re-spacing of the total information Q at locations r_j.

To complete the solution of the problem, it remains to show that $-1 \leqq r_j \leqq 1$. Suppose that the r_j are such that two or more of the r_j are not in the closed interval $[-1, 1]$. Say that r_1 and r_2 are not in this interval. Since c_{12} in (3.2) must be zero, it follows that

$$\textstyle\sum_\epsilon w_\epsilon(x_\epsilon - r_1)(x_\epsilon - r_2)(x_\epsilon - r_3)^2 \cdots (x_\epsilon - r_m)^2(x_\epsilon - r_{m+1})^2 = 0. \tag{3.10}$$

Consider that $(x_\epsilon - r_1)(x_\epsilon - r_2)$ never equals zero and always must have the same algebraic sign. Furthermore

$$(x_\epsilon - r_3)^2 \cdots (x_\epsilon - r_m)^2(x_\epsilon - r_{m+1})^2 \geqq 0. \tag{3.11}$$

Hence equality to zero in (3.10) implies that all terms must be zero. But

$$(x_\epsilon - r_1)(x_\epsilon - r_2)$$

can never vanish, and (3.11) can vanish for at most $(m - 1)$ distinct x_ϵ. Since there are at least $(m + 2)$ distinct x_ϵ, all terms in (3.10) cannot vanish. Thus, two or more of the r_j cannot be excluded from the closed interval $[-1, 1]$, and hence, it has been shown that m of the r_j are in the closed interval $[-1, 1]$.

Consider now the polynomial in (3.6) whose roots are the r_j. In determinant form this polynomial may be shown to be

$$\mathcal{P}(r) = \frac{(-1)^{m+1}}{\Delta} \begin{vmatrix} 1 & r & \cdots r^m & r^{m+1} \\ f_0 & f_1 & \cdots f_m & f_{m+1} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ f_m & f_{m+1} & \cdots f_{2m} & f_{2m+1} \end{vmatrix}, \tag{3.12}$$

where $\Delta = | X'WX |$.

Evaluate $\mathcal{P}(r)$ at $r = -1$ and $r = 1$. It will be seen that, for $J = -1, 1$,

$$(3.13)\qquad \mathcal{P}(J) = \frac{J^{m+1}}{\Delta}\begin{vmatrix} f_0 - Jf_1 & f_1 - Jf_2 & \cdots & f_m - Jf_{m+1} \\ f_1 - Jf_2 & f_2 - Jf_3 & \cdots & f_{m+1} - Jf_{m+2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ f_m - Jf_{m+1} & f_{m+1} - Jf_{m+2} & \cdots & f_{2m} - Jf_{2m+1} \end{vmatrix}.$$

Remembering the definition of f_L, the elements of the indicated determinant $|H_J|$ in (3.13) are of the form $\sum_\epsilon w_\epsilon(1 - Jx_\epsilon)x_\epsilon^L$, $L = O(1)2m$, which shows that

$$(3.14)\qquad H_J = X'W_JX,$$

where W_J is the $N \times N$ diagonal matrix with diagonal elements $w_\epsilon(1 - Jx_\epsilon)$. Since $\min x_\epsilon = -1$ and $\max x_\epsilon = 1$, it follows that W_J always has one diagonal element equal to zero, all others being positive. Hence, from (3.14) it follows that $H_J = X_J'V_JX_J$, where V_J is the $(N - 1) \times (N - 1)$ diagonal matrix formed from W_J by striking out the row and column corresponding to $\min x_\epsilon$ for $J = -1$ and $\max x_\epsilon$ for $J = 1$; and X_J is the $(N - 1) \times (m + 1)$ matrix formed from X by striking out the row corresponding to $\min x_\epsilon$ for $J = -1$ and $\max x_\epsilon$ for $J = 1$. Since there are at least $(m + 2)$ distinct x_ϵ, X_J has rank $(m + 1)$. Also, since V_J is diagonal with nonzero diagonal elements, V_J is nonsingular. It follows that the characteristic numbers of H_J are all positive and hence, $|H_J| > 0$. Then, since $\Delta > 0$, the following conclusions can now be made concerning $\mathcal{P}(J)$ in (3.13):

$$(3.15)\qquad \begin{aligned} &\mathcal{P}(1) > 0, \text{ and } \mathcal{P}(\infty) > 0, \\ &\mathcal{P}(-1) > 0 \text{ and } \mathcal{P}(-\infty) > 0 \text{ for odd } m, \\ &\mathcal{P}(-1) < 0 \text{ and } \mathcal{P}(-\infty) < 0 \text{ for even } m. \end{aligned}$$

It was previously shown that m of the roots r_j are in the closed interval $[-1, 1]$. (3.15) shows that -1 and 1 cannot be roots, and hence, it may be stated that m of the roots are in the open interval $(-1, 1)$. Furthermore, knowing the sign of $\mathcal{P}(r)$ for $r = -1, 1$ and for sufficiently large values of $|r|$, it may be reasoned that all r_j are in the open interval $(-1, 1)$ for one exterior root would imply another.

In conclusion, it has been shown that for $N > (m + 1)$, $-1 < r_j < 1$. As indicated in the preliminary discussion, for $N = (m + 1)$, $-1 \leqq r_j \leqq 1$. Hence, $-1 \leqq r_j \leqq 1$ holds for all cases, and the solution of the problem is complete.

In summary, the r_j are the roots of the polynomial (3.6); the polynomial coefficients β_j satisfy the linear system (3.7). The information located at each r_j is given by (3.9).

Two closing remarks are in order. Returning to (3.2), it may be noticed that the constraints (3.3) are sufficient to make U a diagonal matrix. The added constraint (3.4) is sufficient to locate all r_j in the closed interval $[-1, 1]$. To see this, suppose (3.4) is not imposed. This is equivalent to striking out the last row of $X'WX$ in (3.7), leaving a linear system of m equations in $(m + 1)$ unknowns. The rank of this system is m, and hence, β_{m+1} can be chosen at will. Now, U being diagonal, that is, (3.2) vanishing, demands that m of the r_j be in the closed interval $[-1, 1]$. Accordingly, since $\beta_{m+1} = -\sum r_j$, by choosing $|\beta_{m+1}|$ sufficiently large, a root can always be made exterior to $[-1, 1]$.

It is of further interest to note that the results of this paper suggest an optimum spacing characteristic; namely: $\max r_j - \min r_j \leqq \max x_\epsilon - \min x_\epsilon$, the equality being necessary only for the trivial case $N = (m + 1)$. Thus, without triviality, the same information matrix can be attained by a lesser number of locations in a shorter interval.

4. An application. The application of the results of this paper to the problems listed in Section 1 will be indicated by considering an interval interpolation problem for the quadratic.

Let $m = 2$, and permit N independent observations $y(x_\epsilon)$, $\epsilon = 1(1)N$, of equal variance σ^2 to be taken in the specified interval $x_L \leqq x_\epsilon \leqq x_H$. Let $Y(\xi)$ be the least squares estimator of $P(\xi)$. The problem is to find the spacing of the N observations that will minimize the maximum variance of $Y(\xi)$ for $x_L \leqq \xi \leqq x_H$.

The variance of $Y(\xi)$ is $\sigma^2 Y(\xi) = \vec{\xi}(X'WX)^{-1}\vec{\xi}'$. Whatever the optimum spacing be, it will give rise to some matrix $X'WX$; let this be $(X'WX)_0$. From the given results, it follows that there exists a matrix $R'UR = (X'WX)_0$. Since $m = 2$,

$$R = \begin{pmatrix} 1 & r_1 & r_1^2 \\ 1 & r_2 & r_2^2 \\ 1 & r_3 & r_3^2 \end{pmatrix} \quad \text{and } U = \begin{pmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{pmatrix} \frac{1}{\sigma^2},$$

implying n_j observations at r_j satisfying $x_L \leqq r_j \leqq x_H$, and $\sum n_j = N$, $j = 1(1)3$. Hence, three locations suffice to establish the desired optimum spacing.

Let $\bar{y}_j$ be the average of the n_j observations at r_j; let $\vec{\eta}$ be the column vector $(\bar{y}_1 \bar{y}_2 \bar{y}_3)'$; and let $\vec{a}$ be the column satisfying the normal equations $R'UR\vec{a} = R'U\vec{\eta}$. Since R and U are nonsingular, $R\vec{a} = \vec{\eta}$. Since $R\vec{a}$ is the column vector $(Y(r_1)\ Y(r_2)\ Y(r_3))'$, $Y(\xi)$ passes through $\bar{y}_j$ at $\xi = r_j$. Hence, $Y(\xi)$ may be written in the Lagrange form of Section 2,

$$(4.1)\quad Y(\xi) = \frac{(\xi - r_2)(\xi - r_3)}{(r_1 - r_2)(r_1 - r_3)}\bar{y}_1 + \frac{(\xi - r_1)(\xi - r_3)}{(r_2 - r_1)(r_2 - r_3)}\bar{y}_2 + \frac{(\xi - r_1)(\xi - r_2)}{(r_3 - r_1)(r_3 - r_2)}\bar{y}_3.$$

It follows that $\sigma^2 Y(r_j) = \sigma^2/n_j$. For any such spacing, let $\sigma^2_{\max}$ be the maximum variance of $Y(\xi)$ in the interval. Then,

$$\sigma^2/n_j \leqq \sigma^2_{\max},$$

and thus,

$$\frac{\sigma^2}{3}\sum_1^3 \frac{1}{n_j} \leqq \sigma^2_{\max}. \tag{4.2}$$

The minimum value of $\sum_1^3 1/n_j$ constrained by $\sum_1^3 n_j = N$ is found to be $9/N$ with $n_j = N/3$. Hence, from (4.2),

$$3\sigma^2/N \leqq \min \sigma^2_{\max}. \tag{4.3}$$

Note from (4.1) that $\sigma^2 Y(\xi)$ increases as ξ departs from the smallest and the largest r_j in direction of leaving the interval x_L, x_H. Hence, locate $N/3$ observations at x_L and x_H. Note that $\sigma^2 Y(\xi)$ has one differentiable maximum occurring in the interior of the interval. Locate $N/3$ observations at $(x_L + x_H)/2$. From symmetry, the differentiable maximum then occurs at $(x_L + x_H)/2$. Hence, the maximum $\sigma^2 Y(\xi)$ for $x_L \leqq \xi \leqq x_H$ for the spacing $r_1 = x_L$, $r_2 = (x_L + x_H)/2$, $r_3 = x_H$, $n_j = N/3$ is $3\sigma^2/N$. The inequality (4.3) assures that this particular spacing gives the desired minimum maximum variance, and that this variance is $3\sigma^2/N$, for the equality has been produced.

This conclusion is directly applicable for N divisible by 3. For small N not divisible by 3, a fine structure study, using $3\sigma^2/N$ as basis for comparison, will indicate an acceptable spacing with little increase in variance.