J. Inst. Maths Applics (1973) **12,** *97-102*

The Error of a Function Approximation Based on Random Selected Points

M. MEZEI†

United Hungarian Chemical Works, Budapest, Hungary
[Received 16 June, 1972 and in revised form 1 November 1972]

† Present address: 14 Clubway, Hartsdale N.Y. 10530, U.S.A.

The advantages of the Monte-Carlo quadrature for solving least-square approximation problems are discussed, an approximation algorithm and a probabilistic error estimate are given.

1. Introduction

THE PROBLEM of replacing a complicated function with a simple one — having approximately the same properties often arises both in mathematics and physics. Let us suppose that we have an algorithm for computing the value of the complicated function at any chosen point and we do not have a good estimate about the number and place of points where the value of the complicated function is to be computed in order to get the simple one which may replace it. In this case a random point selection appears advantageous and we give an estimate for the error of the approximation in the cases where this kind of point selection is applied.

2. The "Goodness" Criterion of an Approximation

Let us suppose that we have a complicated function $q(\mathbf{x}), \mathbf{x} = (\mathbf{x}_1, \ldots, x_n)$ which is bounded in a closed finite domain B of the n dimensional Euclidean space, with volume V and the function $p(\mathbf{x})$ is regarded as an approximation to $q(\mathbf{x})$. The difference between the two functions can be characterized in several ways. For example, we can choose some functional $D(p(\mathbf{x}))$ whose value will be the number characterizing the difference. In this paper we will first restrict ourself to the case when

$$D(p(\mathbf{x})) = \frac{1}{V}\int_B (q(\mathbf{x}) - p(\mathbf{x}))^2 dv. \tag{1}$$

3. The Quadrature Problem

As we chose the functional defined by equation (1), we have a numerical quadrature problem. Quadratures other than the Monte-Carlo quadratures have the following drawbacks: (a) In case of equidistant point selection we have to increase the number of points at the rate k^n (k is the number of points in one direction, n is the number of variables); (b) In case of polynomial quadratures — beside the drawback mentioned in (a) — when we want to increase the number of points we often can not use the points used before. On the other hand the Monte-Carlo quadrature leads to an error greater than those mentioned above. As we are computing the integral only to obtain an "approximating" function $p(\mathbf{x})$, an error in the integral value will have as result that $p(\mathbf{x})$ will not be the best approximation to $q(\mathbf{x})$ (in the least square sense).

It is not very serious, however, because: (a) We will have a statistical upper and lower bound for the exact integral value; (b) The error is only second order compared with the error committed by replacing $q(\mathbf{x})$ with $p(\mathbf{x})$.

Consequently, the Monte-Carlo quadrature appears to be a good choice in the following cases: (a) A solution, not far from the best approximation, but certainly not the best approximation is sufficient for us; (b) Because of the unknown nature of the problem the possibility of increasing the number of points used in the approximation several times cannot be excluded; (c) The function $q(\mathbf{x})$ to be approximated lies or nearly lies in the function space T, defined by the parameters of $p(\mathbf{x})$, that is the scalar product of $q(\mathbf{x})$ with its projection to T is near to 1. (If the number of parameters in the function $p(\mathbf{x})$ is n and $q(\mathbf{x})$ lies in T then n independent points are already sufficient.)

4. The Error Estimate

As the function $p(\mathbf{x}))$ contains some parameters which define a function space T, the approximation problem can be formulated as a search for a function $p_o(\mathbf{x})$ which satisfies the following equation:

$$\min_{p(\mathbf{x})\in T} D(p(\mathbf{x})) = D(p_o(\mathbf{x})) \tag{2}$$

where D is defined by equation (1).

If the values $(q(\mathbf{x}_i) - p_1(\mathbf{x}_j))^2$ are regarded as values of a probability variable, their expectation value will be $D(p(\mathbf{x}))$. Let A and C be two sets of n_A and n_C points, respectively, whose elements are equally distributed in B, $p_1(\mathbf{x})$ be a function that

$$\frac{1}{n_A} = \sum_{\mathbf{x}_i\in A} (q(\mathbf{x}_i) - p_1(\mathbf{x}_i))^2 \tag{3}$$

is minimal ($p_1(\mathbf{x}) \in T$). Then we can prove the following inequalities which can serve as an error estimate:

$$M(\frac{1}{n_A}(q(\mathbf{x}_i) - p_i(\mathbf{x}))^2) \leq D(p_o(\mathbf{x})) \leq M(\frac{1}{n_C}\sum_{\mathbf{x}_i\in C}(q(\mathbf{x}_i) - p_1(\mathbf{x}_i))^2), \tag{4}$$

where $M(\ )$ means the expectation value.

As the points $\mathbf{x}_i$ are equally distributed in B, it follows that

$$M(\frac{1}{n_C}\sum_{\mathbf{x}_i\in C}(q(\mathbf{x}_i) - p_1(\mathbf{x}_i))^2) = D(p_1(\mathbf{x})). \tag{5}$$

The right-hand inequality follows from equation (5), because $D(p_o(\mathbf{x}))$ is the minimal value of the functional D in the function space T and $p_1(\mathbf{x}) \in T$. Moreover, from the definition of $p_1(\mathbf{x})$ it follows that

$$\frac{1}{n_A}\sum_{\mathbf{x}_i\in A}(q(\mathbf{x}_i) - p_1(\mathbf{x}_i))^2 \leq \frac{1}{n_A}\sum_{\mathbf{x}_i\in A}(q(\mathbf{x}_i) - p_o(\mathbf{x}_i))^2 \tag{6}$$

and also for the expectation values:

$$M(\frac{1}{n_A}\sum_{\mathbf{x}_i\in A}(q(\mathbf{x}_i) - p_1(\mathbf{x}_i))^2) \leq M(\frac{1}{n_A}\sum_{\mathbf{x}_i\in A}(q(\mathbf{x}_i) - p_o(\mathbf{x}_i))^2). \tag{7}$$

As the points $\mathbf{x}_i$ are equally distributed in B:

$$M(\frac{1}{n_A}\sum_{\mathbf{x}_i \in A}(q(\mathbf{x}_i) - p_o(\mathbf{x}_i))^2) = D(p_1(\mathbf{x})). \tag{8}$$

Putting equation (8) into equation (7) we obtain the left-hand inequality of (4).

4.1. *Remarks*

(i) From the proof of the left-hand inequality it follows that if $p_i(\mathbf{x})$ minimizes D then the equality holds in the left-hand inequality of (4).

(ii) From the definition of $p_o(\mathbf{x})$ and $p_i(\mathbf{x})$ it follows that $p_1(\mathbf{x})$ minimizes D if and only if the equality holds in the right-hand inequality of (4).

(iii) From the first two remarks and from the inequalities (4) it follows that $p_1(\mathbf{x})$ minimizes D if and only if

$$M(\frac{1}{n_A}\sum_{\mathbf{x}_i \in A}(q(\mathbf{x}_i) - p_1(\mathbf{x}_i))^2) \leq M(\frac{1}{n_C}\sum_{\mathbf{x}_i \in C}(q(\mathbf{x}_i) - p_1(\mathbf{x}_i))^2). \tag{9}$$

(iv) As the choice of the points of A and C did not depend on the function(s) minimizing D, the distributions of the values $(q(\mathbf{x}_i) - p_o(\mathbf{x}_i))^2$ will be the same over both sets.

According to remark (iii), if equation (9) holds, then $p_i(\mathbf{x}))$ is identical with one of the functions minimizing D (for example $p_o(\mathbf{x})$) and therefore also the distributions of the values $(q(\mathbf{x}_i) - p_1(\mathbf{x}_i))^2$ must agree over both sets, that is the identity of the two distributions is also a necessary and sufficient condition for $p_1(\mathbf{x})$ to minimize D.

(v) An inclusion theorem for the Tschebyscheff approximation, showing some similarity to our error estimate was proven by Collatz (Garabedian, 1965). This theorem provided the points $\mathbf{x}_i$ satisfy a system of inequalities gives upper and lower bounds to $D(p_o(\mathbf{x}))$ which can be analytically expressed from the $p_1(\mathbf{x})$.

5. The Extension of the Functional $D(p(\mathbf{x}))$

As can be seen in the proof of inequalities (4), any functional $D'(p(\mathbf{x}))$ having the form

$$D'(p(\mathbf{x})) = \frac{1}{V}\int_B f(p(\mathbf{x}))dv. \tag{10}$$

where $f(u)$ is an integrable function in B, satisfies them. Thus our estimate can be used for several other "goodness criterions" too. An interesting example is when the functional to be minimized is

$$D'(p(\mathbf{x})) = \frac{1}{V}\int_B \rho(\mathbf{x})(q(\mathbf{x}) - p(\mathbf{x}))^2 dv \tag{11}$$

where $\rho(\mathbf{x})$ satisfies the following equations:

$$\rho(\mathbf{x}) > 0 \tag{12}$$

$$\int_B \rho(\mathbf{x})dv = 1 \tag{13}$$

that is $\rho(\mathbf{x})$ is a weighting function.

This function can be handled in a very simple way: It is known from the theory of Monte-Carlo integration (Shreider, 1966) that if the points $\mathbf{x}_i$ are distributed according to a probability distribution having $\rho(\mathbf{x}))$ as its density function, then

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{f(\mathbf{x}_i)}{\rho(\mathbf{x}_i)} = \int_B f(\mathbf{x})dv. \tag{14}$$

When replacing $f(\mathbf{x})_i)$ by $\rho(\mathbf{x}_i)(q(\mathbf{x}_i)-p(\mathbf{x}_i))^2$ we can see that if the points of the sets A and C are distributed according to $\rho(\mathbf{x})$ (which is possible because the equations (12) and (13) ensure that $\rho(\mathbf{x})$ can be a density function), the expression to be minimized will have similar form as in the case when only $(q(\mathbf{x})-p(\mathbf{x}))^2$ was to be integrated so the same approximation algorithm (see Section 7) can be applied for both cases.

6. The Use of the Inequalities (4)

These inequalities can serve the following aims: (i) They give a statistical estimate to the upper and lower bounds of the error of the best approximation; (ii) Since equation (9) is a necessary and sufficient condition for $p_1(\mathbf{x})$ to minimize D, we also have a criterion to decide the refinement of the approximation. As equation (9) contains expectational values, whose exact value is not known, only their statistical estimates, the validity of equation (9) can be checked only by statistical methods.

The Student t-test, for example, is appropriate for this reason, because, according to remark (iv), if a function $p(\mathbf{x}))$ minimizes D then not only the expectational values of $(q(\mathbf{x}_j) - p(\mathbf{x}_i))^2$, but their distributions too must agree in A and C, so the applicability of the Student t-test (the equality of variances) is ensured. It must be pointed out, however, that the equality of variances is only a necessary condition for $p(\mathbf{x})$ to minimize D.

7. The Algorithm

Let us suppose that the $p_1(\mathbf{x})$ minimizing the expression (3) is calculated by the conditions that the partial derivatives of (3) with respect to the parameters of $p_1(\mathbf{x})$ must vanish. As the derivation is a linear operator, we can see at once that the equations determining the wanted parameter values will be of the form:

$$\frac{1}{n_A} = \sum_{\mathbf{x}_i \in A} (q(\mathbf{x}_i) - p_1(\mathbf{x}_i))2 \qquad (15)$$

where $(\ )'$ denotes the derivate of the expression $(\ )$ with respect to one parameter. (It can be noticed that every member of the sum depends only on one point $\mathbf{x}_i$.)

The algorithm will then consist of the following steps:

(i) Select from B the points for A, calculate the corresponding $q(\mathbf{x}_i)$ values and prepare the derivatives appearing in equation (15) for $\mathbf{x}_i \in A$.

(ii) Select from B the points for C, calculate the corresponding $q(\mathbf{x}_i)$ values and prepare the derivatives appearing in equation (15) for $\mathbf{x}_i \in C$.

(iii) Solve the equation (15) corresponding to A. If the last test (step vi) permits to finish the computation then stop.

(iv) Compute the deviations $(q(\mathbf{x}_i) - p_1(\mathbf{x}_i))^2$ for $\mathbf{x}_i \in A$ and $\mathbf{x}_i \in C$, respectively, and compute the values of the statistical functions t and F.

(v) Form a "new" set A by merging the points $\mathbf{x}_i \in C$ into the former set A, and form a new equation (15) according to the new A (in case of linear parameters simply add the equation (15) resp. C to the equation (15) resp. former A).

(vi) If both the t and F test do not show significant difference (the level of significance will be characteristic to the exactness of the approximation) then go to step (iii) otherwise go to step (ii).

7.1. *Remarks on the Algorithm*

(i) It must be pointed out that we have to solve our equation (15) exactly in every iteration to be sure that the inequalities (4) on which the end criterion is based, hold. This is not very serious in general, because (at least from the second iteration) we have a good starting solution (i.e. the solution obtained in step (iii) of the previous iteration).

(ii) In the algorithm, beside the solving of equation (15) we have to check the equality of two expectational values and/or the identity of two distributions. The present algorithm uses the t and F test for this purpose. It must be pointed out, however, that at this point other statistical methods can be used as well.

8. Numerical Example

To show an example, we approximated the function e^{-x} in the [0,1] interval by polynomials of different degrees. The results are collected in Tables 1-3. D_A and D_C are

$$\frac{1}{n_A} = \sum_{\mathbf{x}_i \in A} (q(\mathbf{x}_i) - p_1(\mathbf{x}_i))2 \quad \text{and} \quad \frac{1}{n_C} = \sum_{\mathbf{x}_i \in C} (q(\mathbf{x}_i) - p_1(\mathbf{x}_i))2$$

respectively, n is the degree of the approximating polynomial.

TABLE 1

$n = 1$

$D(p_o(\mathbf{x})) = 5.3_{10}{-4}$

n_A	n_C	D_A	D_C	t	F
20	20	4.5104	6.5104	1.1	1.6
40	20	5.3104	6.8104	14	.1.2
60	20	5.6104	6.2104	0.33	1.6
80	20	5.7104	4.5104	0.70	2.6

TABLE 2

$n = 2$

$D(p_o(\mathbf{x})) = 3.7_{10}{-6}$

n_A	n_C	D_A	D_C	t	F
20	20	3.3106	4.4106	0.74	2.2
40	20	3.6106	4.3106	0.65	2.3
60	20	3.810-6	5.2i06	1.3	2.2
80	20	4.1106	2.9106	1.3	2.8

TABLE 3

$n = 3$

$D(p_o(\mathbf{x})) = 1.4_{10}{-8}$

n_A	n_C	D_A	D_C	t	F
20	20	1.1108	1.4108	0.6	2.2
40	20	1.2108	1.5108	0.6	5.1
60	20	1.2108	1.4108	0.39	5.2
80	20	1.3108	1.41a8	0.45	1.7

The degrees of freedom corresponding to a t and an F value can be obtained from the corresponding n_A and n_C values: that for the t value is $n_A + n_C - 1$ while those for the F value are n_A and n_C themselves, n_A corresponding to the denominator and n_C to the numerator.

From the results it seems that:

(i) The D_A and D_C values are not very different and in most of the cases the $D(p_o(\mathbf{x})))$ value is bounded by them. The "anomalies" occur in the cases where the number of elements in A and C are very different and the approximation is poorer. In any case, however, their values are not far from the $D(p_o(\mathbf{x})))$ value.

(ii) The value of t becomes small much easier and does not show as large oscillations as the F value. As the t test requires a small F value, it seems very advisable to study other statistical tests as well.

In addition to this simple example it must be remarked that the approximation algorithm described in Section 7 was previously applied with success in a quantum chemical approximation problem (Mezei, 1972).

Thanks are due to Katalin Bencsath for her important remarks on the manuscript.

REFERENCES

GARABEDIAN, H. L. (ed.) 1965 *Approximation of functions.* Elsevier.

MEZEI, M. 1972 *Lokális energia módszerek* (Local-Energy Methods). Ph.D. Thesis, Eötvös Loránd University, Budapest.

SHREIDER, JU. A. (ed.) 1966 *The Monte Carlo method.* Oxford: Pergamon.