STATISTICALLY BASED INTERPOLATION AND EXTRAPOLATION IN AUTOMATED TESTING SYSTEMS*

Frederic H. Herman and Thaddeus J. Kobylarz
Systems Design and Analysis Co.
West New York, New Jersey

Abstract

This paper addresses itself to the problem of automatically testing systems which are representable by continuous, instantaneous functions. It is assumed that the tests are controlled and analyzed by a digital computer. Since discrete measurements are required, interpolation and extrapolation must be considered. In order to circumvent the deficiencies of conventional methods employing deterministic approaches, statistically based interpolation and extrapolation methods are described.

1. INTRODUCTION

Automated testing of systems, which are representable by instantaneous continuous functions, is typically controlled by a digital computer. In addition to controlling the test, the results are analyzed by the computer to determine if the input/output function satisfies predetermined constraints of acceptability.

A disparity exists in that the systems being tested are continuous and the systems performing the tests are discrete. A number of important problems arise from the disparity.(1) For example, it is obvious that exhaustive testing is not possible. Economy closely governs the number and location of test points allowable for a given test. A certain amount of time is required for the conversion of data between analogue and digital forms. Furthermore, testing may involve non-electrical inputs such as temperature, pressure, and humidity. Outputs may also be non-electrical. The testing system may be limited in the range of inputs that it can simulate. Field-testing of systems may require that the automated tester be far less complex than its laboratory counterpart.

Since the response function must be evaluated between the discrete test points, it is apparent that interpolation and extrapolation are vital considerations for such automated testing systems.

Deterministic approaches, such as power series curve-fitting, have a serious deficiency. While methods such as least squares curve-fitting, enable the average error to be monotonically reduced as the order of the fit is increased,(2) absolute error bounds are either impossible to obtain or are so crude as to render them valueless.(3) The "Bernstein polynomial"(4) is an example of error bounding methods in which it can be assumed that the response function is uniformly continuous. This method yields an extremely pessimistic measure of the error.

In order to circumvent the deficiencies of conventional methods, a probabilistic approach to interpolation and extrapolation is taken. It is assumed that the systems to be tested come from a large population and that it is possible to obtain the probability distribution of the input/output derivative as a function of the input. Thus a "learning period" is required in which many devices are tested much more thoroughly than is economically feasible in the field. Due to the averaging characteristic of statistics, the distribution itself will vary more "smoothly" over the input range than the input/output response function of any particular system. For instance, the density function(5) may be found to be Guassian-normal(5) in which the mean and the variance are functions of the input. In this paper we consider how such information can be used to automatically test additional systems, for which economically constrained testing exists.

*This work was supported by the Engineering Sciences Division, Picatinny Arsenal, under Contract DAAA21-72-C-0506

2. BASIC TECHNIQUE

Consider a system having an input/output function $f(x)$ which is continuous, time-invariant, but allowably non-linear. Assume that the probability distribution of the input/output derivative $df(x)/dx$ is, in general, x dependent, i. e.

$$p\{a \leq y'(x) \leq b\} \equiv \int_a^b h\ (x,z)dz. \tag{1}$$

where: $y'(x) \equiv \frac{df(x)}{dx}$.

An example of such statistical knowledge can correspond to a normally (Gaussian-normal) distributed derivative, having a mean and variance as some known function of x.

At some x_o, the system is tested for its response $y(x_o)$. We wish to combine the knowledge of this test result with our previous statistical information. From the requirement of $f(x)$ continuous, there is some perturbation Δx for which $f(x)$ is nearly linear in the interval $(x_o, x_o + \Delta x)$. (One may verify this by considering the Taylor expansion of $f(x)$ about x_o.) Therefore, for this perturbation, we have

$$y(x_o+\Delta x) \approx y(x_o) + \Delta x \cdot y'(x_o). \tag{2}$$

Let us choose some confidence interval for the random variable $y'(x_o)$; for some confidence value α:

$$\alpha = p\{a \leq y'(x_o) \leq b\} = \int_a^b h(x_o,z)dz. \tag{3}$$

For a non-symmetric density function, it is somewhat arbitrary how one picks the confidence interval (C. I.) since the mode and the mean do not generally coincide. In order to simplify what follows, we assume symmetry and take the C. I. about the mean of the distribution.

Using the C. I. boundary in equation (3), a C. I. is obviously established for $y(x_o+\Delta x)$. Thus,

$$\alpha = p\{y(x_o)+a\cdot\Delta x \leq y(x_o+\Delta x) \leq y(x_o)+b\cdot\Delta x\}. \tag{4}$$

For a negative excursion, assuming the same Δx is appropriate, one has

$$\alpha = p\{y(x_o)-b\cdot\Delta x \leq y(x_o-\Delta x) \leq y(x_o)-a\cdot\Delta x\}. \tag{5}$$

Figure 1 depicts equations (4) and (5).

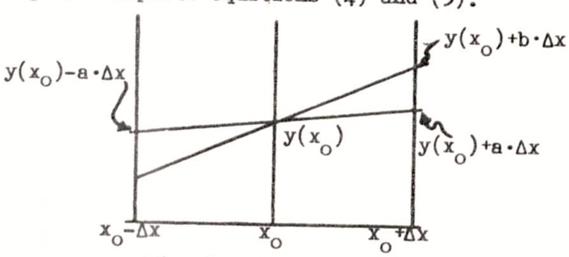

Fig. 1 Simple extrapolation.

The above basic process allows extrapolation of confidence bounds from a point of measurement corresponding to deterministic data. The statistically based extrapolation introduces a C. I., not occuring in conventional methods. Additionally, some perturbation error is present from the approximate nature of (2). It is assumed that Δx is taken small enough to make this error negligible. One may easily extend the simple process described above to an interpolation scheme. Once the single-point extrapolation has been completed, a second increment is chosen appropriate for equation (2). Denote it as Δx_1, indicating that it was chosen at $x_1 = x_o+\Delta x_o+\Delta x_1$. The system is tested for its response at x_1 and then single-point extrapolation is performed at the point x_1. Figure 2 illustrates the result.

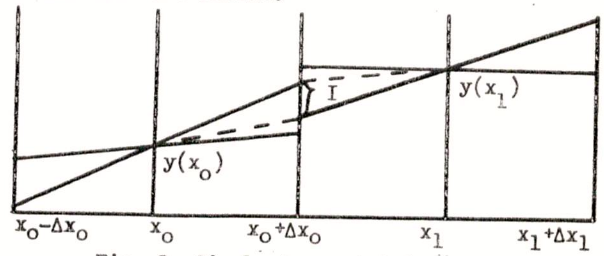

Fig. 2 Simple two-point interpolation.

In general the two bounds at $x_o+\Delta x_o$, resulting from x_1 and x_2, will not coincide but will have some region of intersection, if the confidence level is large ($\alpha > .95$). Lack of intersection means that at least one of the random variables, $y(x_o)$ and $y(x_1)$, was actually outside its corresponding C. I. Additional testing is required at $x_o+\Delta x_o$ for each case of lack of intersection, but such an occurance is only $(1-\alpha^2)$ probable. For $\alpha = .96$, the probability of non-intersection is .0784 or 7.84%. This computation assumes independence of the two bounds. If an intersection exists, one may use the intersection as the composite C. I. with probability α^2. The overall bounds on $f(x)$ in the interval (x_o,x_1) then is formed by reducing the C. I.'s for each random variable until they both coincide with the intersection, I (see Figure 2).

The above interpolation scheme can be continued to the right and to the left to include any overall increment. Between deterministic test points, $f(x)$ is bounded. This bound is piecewise linear and can be represented as two piecewise linear functions.

3. EXTRAPOLATION BY ITERATION

It is obvious that even the above scheme will often require an enormous number of test measurements. When Δx must be kept very small, the simple statistical based interpolation and extrapolation may offer only a small advantage over conventional means. The statistical information will now be used to fuller advantage.

Consider the following construction. Referring to Figure 1, at $x_o+\Delta x_o$, a C. I. has been established $(y(x_o)+a\cdot\Delta x_o,\ y(x_o)+b\cdot\Delta x_o)$. At each of the C. I. boundaries, construct a C. I. as if each of these points were actually test measurement data. Call the point $x_o+\Delta x_o, x_1$. Using these two C. I.'s, a C. I. can be constructed which corresponds to the value $y(x_1+\Delta x_1)$. The maximum of this C. I. is simply the maximum of the C. I. constructed from $(x_1, y(x_o)+b\cdot\Delta x_1)$. Similarly, the minimum of the C. I. for $y(x_1+\Delta x_1)$ is the minimum boundary of the C. I. constructed at $(x_1, y(x_o)+a\cdot\Delta x_1)$. Figure 3 illustrates the process in which two iterations have been executed. Note that this could be done bilaterally at x_o.

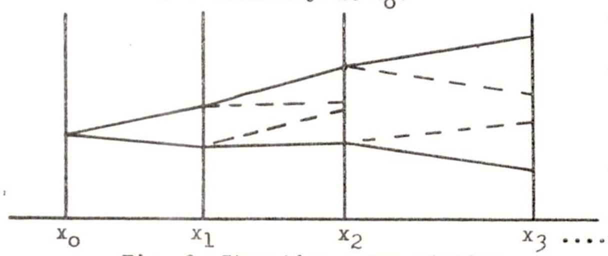

Fig. 3 Iterative extrapolation.

Also it is apparent that the bounds thus formed result in a diverging overall bounds on $y(x)$. As we wish to consider differential intervals shortly, we must be assured that the rate of divergence is not a function of the number of iterations, but of the extrapolation distance, $x-x_o$. It is clear that the density function of population of continuous functions is itself continuous. One only needs to consider the parameters of a distribution. Typically they can be represented as continuous functions of the population values summed over the entire population.(5) Thus they are continuous functions of continuous functions and are therefore themselves continuous. Hence by making Δx small enough, the density function of $y'(x)$ can be made approximately constant over a iteration interval. Inspection of the iterative construction and Figure 3 reveal that for a density function, constant over some interval, the number of iterations used to span the interval has no effect on the iterative C. I.

As the iteration interval becomes a differential distance, the C. I. limits become continuous functions of x. If α is fixed and the C. I. is uniquely defined,* then at each x, these limit functions can be evaluated, such that

$$\alpha = \int_{a(x,\alpha)}^{b(x,\alpha)} h(x,z)dz. \qquad (6)$$

*It was previously assumed that only symmetric distributions would be considered here, and that C. I.'s would be symmetric about the mean.

Moreover, they are derivative functions and can be integrated to obtain the C. I. boundary curves. Thus the upper and lower boundaries of $y(x_r)$ are, respectively

$$y_{u,\alpha}(x_r) = y(x_o) + \int_{x_o}^{x_r} b(x,\alpha)dx = y(x_o) + B(x_r,\alpha) - B(x_o,\alpha)$$

$$y_{L,\alpha}(x_r) = y(x_o) + \int_{x_o}^{x_r} a(x,\alpha)dx = y(x_o) + A(x_r,\alpha) - A(x_o,\alpha). \qquad (7)$$

$$\text{for } x_r \geq x_o.$$

Similarly, for extrapolation to the left of x_o,

$$y_{u,\alpha}(x_1) = y(x_o) + \int_{x_o}^{x_1} a(x,\alpha)dx = y(x_o) + A(x_1,\alpha) - A(x_o,\alpha)$$

$$y_{L,\alpha}(x_1) = y(x_o) + \int_{x_o}^{x_1} b(x,\alpha)dx = y(x_o) + B(x_1,\alpha) - B(x_o,\alpha). \qquad (8)$$

$$\text{for } x_1 \leq x_o.$$

Note that the computations during the actual testing can be minimized if the functions $A(x,\alpha)$ and $B(x,\alpha)$ of (7) and (8) are precomputed for the desired confidence value, α. This is possible since, both of the functions depend only on the distribution of $f'(x)$, the derivative of the system response function.

The construction of iterative extrapolation is intuitively appealing. One feels that the bound so constructed must be at least α probable. This feeling comes from our apparently taking the worst case at each iteration. However, it is not always true that the construction will yield a C. I. of probability α or better. One can formulate situations such as an exponentially distributed derivative using a low value of α, in which it is possible to show that iterative extrapolation does not result in a confidence of at least α. Very briefly, the iterative extrapolation corresponds to a linear combination of random variables. If they can be considered independent, one must examine the n-ary convolution of the distributions. It must be shown that the area under the convolved distribution, between the limits corresponding to the iterative C. I., is greater than or equal to α. Although no general necessary conditions appear to be tractible, it has been shown elsewhere(6) that it is sufficient for the random variables to be normal. Also, many mono-modal symmetric density functions give comparable results.

4. ITERATIVE INTERPOLATION

The iterative construction in the preceding section, can be utilized in an interpolation scheme. This extension proceeds in a similar manner to the extension of the simple extrapolation of Section 2. Observe in Figure 4 that iterative extrapolation was performed at x_0 and x_1. The intersection of the two bounded regions is taken as the composite or interpolated C. I. When no intersection occurs or is very narrow with respect to each of the individual C. I.'s, it is necessary to measure the systems response at an additional point between x_0 and x_1. Similarly, when the composite C. I. is excessive between these two points, additional system measurements are made, in the region that the C. I. was excessive. Also, when some a'priori acceptable region for $f(x)$ is defined, additional measurements may be made when the composite C. I. exceeds this region. Any one of these situations lends itself to an adaptive feature. That is, analysis of the measured data can decide the need and location of additional measurements.

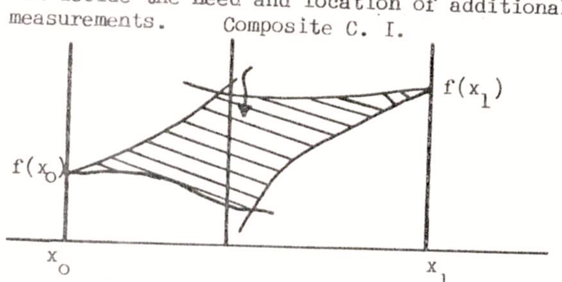

Fig. 4 Iterative interpolation.

5. CONCLUSIONS

The two statistically based interpolation and extrapolation techniques offer some advantage over conventional methods. In testing a system represented by a continuous time invariant response function, it is usually of interest to determine if the function lies within some acceptable bounds. The incorporation of statistical information enables one to avoid the usual deficiencies of conventional extrapolation and interpolation and in particular reduce the burden of the number of measurements on the system.

The major problem with using the statistically based techniques is the requirement of the derivative distribution. While in some physical situations, derivatives are available directly, one may in general require numerical differentiation. This can give rise to severe errors in some cases, so that derivative determination may not be feasible. Even if this problem does not occur, it is apparent that much data must be collected to determine a reasonably good estimate of the derivative distribution. Although the initial investment in obtaining such statistics may be severe, the later savings in testing systems may justify the cost.

A final problem associated with the iterative procedures is that the derivative values at different points are considered independent. Since the statistics are gathered as absolute probabilities, the independence assumption is a naive approach and can be expected to result in a more pessimistic C. I. than if conditional density functions were used. However, the use of absolute density functions simplify the practical aspects of data gathering and computation.

REFERENCES

1. Kobylarz, T., and Herman, F. (1972). "Computerized Performance Evaluation of Instantaneous Systems." Technical Report ESD IR 560, Picatinny Arsenal, Dover, N. J.

2. Mood, A. M., and Graybill, F. A. (1963). Introduction to the Theory of Statistics. McGraw-Hill Book Co., Inc., New York.

3. Hamming, R. W. (1962). Numerical Methods for Scientists and Engineers. McGraw-Hill Book Co., Inc., New York.

4. Ralston, A. (1965). A First Course in Numerical Analysis. McGraw-Hill Book Co., Inc., New York.

5. Feller, W. (1968). An Introduction to Probability Theory and Its Applications. Vol. 1, John Wiley & Sons, Inc., New York.

6. Herman, F., and Kobylarz, T., "Bounds on Linear Combinations of Independent Random Variables". Submitted for publication.