RATIONAL APPROXIMATION IN HIGH SPEED COMPUTING

Cecil Hastings, Jr.

Prepared for presentation at the

IBM Computation Forum

Endicott, N. Y.

P-114

Revised 12-2-49 December 5 thru 9, 1949

The RAND Corporation

· SANTA MONICA · CALIFORNIA

Revised 12-?-49

P-114

SUBJECT: RATIONAL APPROXIMATION IN HIGH SPEED COMPUTING CSM-239

By: Cecil Hastings, Jr.

Copies to:

1. Introduction:

This is a brief report on a study that is being made at RAND on the use of rational approximation in high speed computing. The work we report upon was largely stimulated, in the first place, thru appearance of the IBM Type 604 Calculating Punch, and our work was given further impetus by the reported development of the IBM Card-Programmed Electronic Calculator. The opportunity of doing away with the use of card tables thus presented itself to us, and we began to prepare for the day when compact approximate expressions would take their place in the art of digital computing. The subject of rational approximation then became a matter of increasing importance. We note in passing that proper use of the 604 can eliminate the use of tables to a considerable extent. Thus, to give an example, one can evaluate a fifth (or even higher) degree polynomial in single card computation on a 604. The machine will then read an arbitrary value of 'x' from a detail card and punch out P(x) upon the same card. This capability may be used to compute, for example, 5D sines over the quadrant from such a polynomial expression. Simple rational expressions may also be computed in single card computation. Logarithmns, square roots, and many special functions have been computed directly on the 604 with considerable success. When the Card-Programmed Electronic Calculator is available, our needs for compact approximate rational expressions to univariate functions will be much increased, for in a sequence of calculations we shall often be required to 'look up' a random value of a function in order to continue, and a bulky table is useless here. Some attention is also being given to the problem of multivariate approximation. Our study is largely of an empirical nature. We are compiling data on many instances of rational approximation. A few comments on this data follow. In the remaining sections, a number of random topics pertinent to the study are discussed briefly.

By the term 'primitive approximation' we essentially mean the most accurate approximation that can be achieved in a given set of circumstances. We shall call the graph of

$$(1) \qquad \in(x) = f^*(x) - f(x),$$

the 'error curve' of approximation. The reader will notice that the several error curves displayed in this paper have been truly leveled. That is, our approximations are 'primitive'. Our parameter values are then, of necessity, determined to an excessive number of figures for practical purposes. These may be cut back to obtain 'working' approximations.

1. INTRODUCTION: (Cont'd.)

Each primitive approximation will be described by an accurate error curve, the primitive parameter values, location of roots 'r_i', location of extremal values 'e_i', and the common extremal value 'd'.

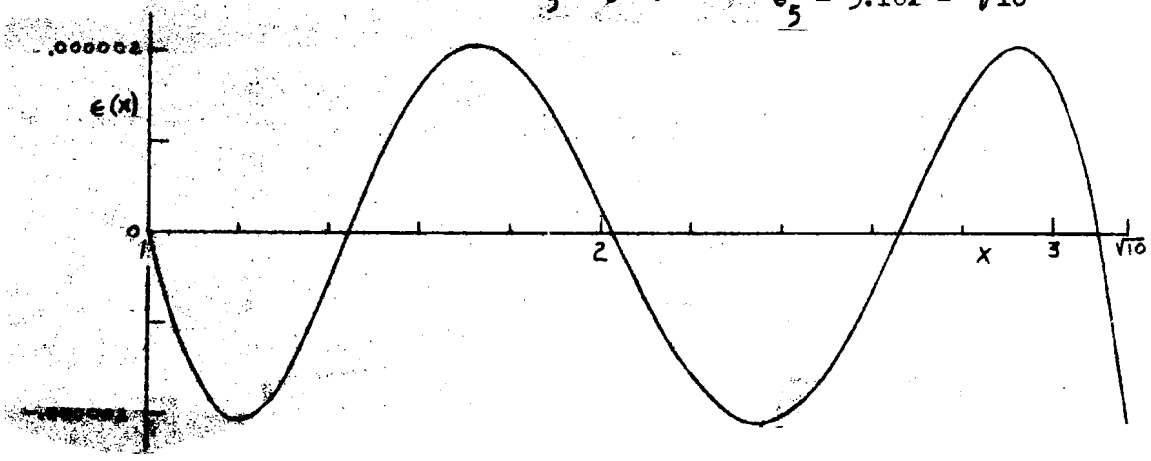
Thus, we approximate the common logarithmic function 'ℓgx' over $(1, \sqrt{10})$ by the form

(2)
$$\xi = \frac{x - 1}{x + 1}$$

$$\ell g^* x = c_1 \xi + c_3 \xi^3 + c_5 \xi^5 + c_7 \xi^7,$$

and record the following data:

$c_1 = .8685,5434$	$r_1 = 1.000$	$e_1 = 1.204$	$d = .0000,0206$
$c_3 = .2911,5068$	$r_2 = 1.446$	$e_2 = 1.722$	
$c_5 = .1536,1371$	$r_3 = 2.028$	$e_3 = 2.348$	
$c_7 = .2113,9497$	$r_4 = 2.656$	$e_4 = 2.920$	
	$r_5 = 3.098$	$e_5 = 3.162 = \sqrt{10}$	



The example is an interesting, and perhaps useful one. We notice that

(3)
$$\eta = \frac{x - \sqrt{10}}{x + \sqrt{10}}$$

$$\ell g^* x = .5 + c_1 \eta + c_3 \eta^3 + c_5 \eta^5 + c_7 \eta^7$$

is an equally good approximation to 'ℓgx' over the full interval (1, 10).

1. Introduction: (Cont'd.)

Each function approximated is studied with respect to some reasonable sequence of forms, and not as an isolated case, lest one gain this impression from the example above. We will thus obtain empirical data concerning rates of convergence, location of roots, location of extremals, and convergence of parameter values.

Reference

C. Lanczos 'Trigonometric Interpolation of Empirical and Analytical Functions' Journal of Mathematics and Physics, Vol. XVII, No. 3, September 1938.

2. The Fitting of Rational Forms: Cont'd.

While the functions approximated have quite different behavior, the error curves are remarkably alike except for scale.

Our first trial in making a rational fit is usually by 'specification of roots'. That is, we impose upon the form those values of 'x' at which the error curve is to cross the x-axis. Then we compute the error curve on an 'evaluation sheet', and if we have a true feeling for the situation, all peaks will be of comparable magnitude.

An evaluation sheet is simply a work sheet containing a list of key values of the function f(x) being approximated and a suitable computing setup for use in evaluation of corresponding values of a given approximation f*(x). A final column allows for the evaluation of $\epsilon(x)$.

To obtain the quality of approximation that we desire to exhibit, an adequate number of accurate function values must be incorporated in the evaluation sheet. Generally, we select about fifty or sixty values of the independent variable in the construction of a given evaluation sheet, and these are given to about three more figures than the most accurate approximation we expect to obtain. With the possible exception of a few extraneous values, our tabulated key values of a function f(x) will be for equally spaced intervals either in 'x' or in some simply related transformed variable $\xi = \xi(x)$.

Once a crude approximation has been obtained and the corresponding error curve determined, we may employ the more refined method of 'specification of extremals' to level the peaks as we have done in the several examples given above. Several computing cycles may be required to achieve the quality of approximation desired.

In speaking of extremal deviations, we naturally refer to the error curve of approximation. Generally an extremal of an error curve will be either a true minimum or a true maximum. Possible exceptions occur only at end points of the interval of interest, in well behaved cases. In general, there will be one more extremal to an error curve than there are parameters free in the form. Thus, the two error curves given above each have four extremals, and there are three parameters (b_0, b_2, b_4) in the form. Or looking at the matter in another light, there are seven extremals, and six parameters, as the approximation holds over $(-\infty, +\infty)$.* Thus a final error curve results from the correct determination of 'n+1' extremal locations 'e_i' or, more helpfully, from the correct determination of 'n' extremal locations 'e_i' and one common deviation 'd', that is the absolute magnitude of the error curve at each of the 'n+1' extremals.

While our problem is then accurately described as one of solving a system of transcendental equation in 'n+1' unknowns 'e_i' and 'd', actually our problem amounts to nothing more than that of solving a single transcendental equation in 'd'. And this is quite a simple matter. In practice, the values

*and odd power terms are missing.

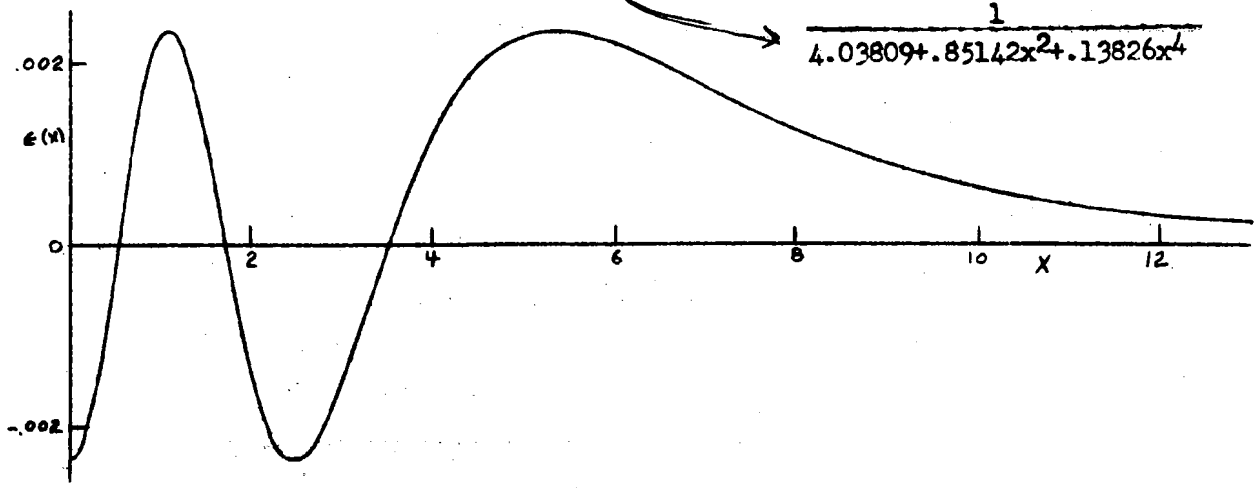
2. The Fitting of Rational Forms:

There are several important ways in which the problem of fitting a rational form may be linearised. One method is that of 'specification of roots'. Another is that of 'specification of extremal deviations'.

We hope to make some useful comments upon the location of roots and extremals. Perhaps our final observations will merely be of an empirical nature. It has become apparent to us that, in comparable situations, one error curve will be much like another of the same order. Thus, consider the examples below. (We shall actually rule out approximations of the type below as being 'non-computable'. Section 3 will throw some light on this aside.)

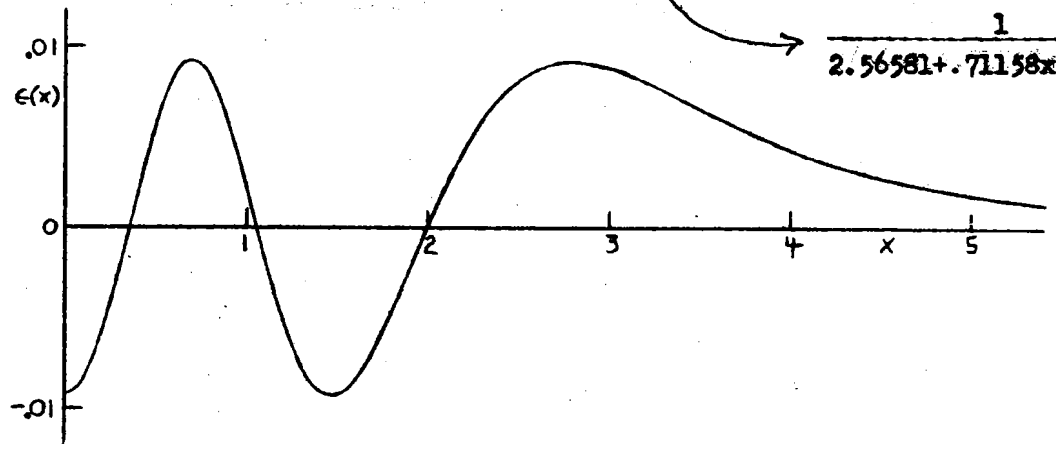
$W(x) = \dfrac{e^{-x}}{(1+e^{-x})^2}$ by

$$\frac{1}{4.03809+.85142x^2+.13826x^4}$$



$E'(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ by

$$\frac{1}{2.56581+.71158x^2+.81788x^4}$$



2. The Fitting of Rational Forms: (Cont'd.)

of 'e_1*' that we 'read' from a given intermediate error curve are used unchanged in the cycle that follows. The drift of location is usually very small, and the results are somewhat insensitive to the 'e_1*'. (An abbreviated numerical process is used in lieu of curve plotting, and hence 'read' is in quotes.)

(3) The Problem of Sensitivity:

In almost all cases of rational approximation, we run into a problem of 'sensitivity'. Thus, a function f(x) may behave decently, have an adequate rational approximation of the form

$$(1) \qquad f^*(x) = \frac{N(x)}{D(x)}$$

and yet the components N(x) and D(x) may behave very badly, changing in joint fashion by many orders of magnitude as 'x' ranges over the interval of approximation. Fortunately, however, there always seems to be a happy solution to this kind of difficulty.

Thus, in making a 4D (d = .000139) approximation to

$$(2) \qquad \phi(x) = \frac{1 - e^{-x}}{x}$$

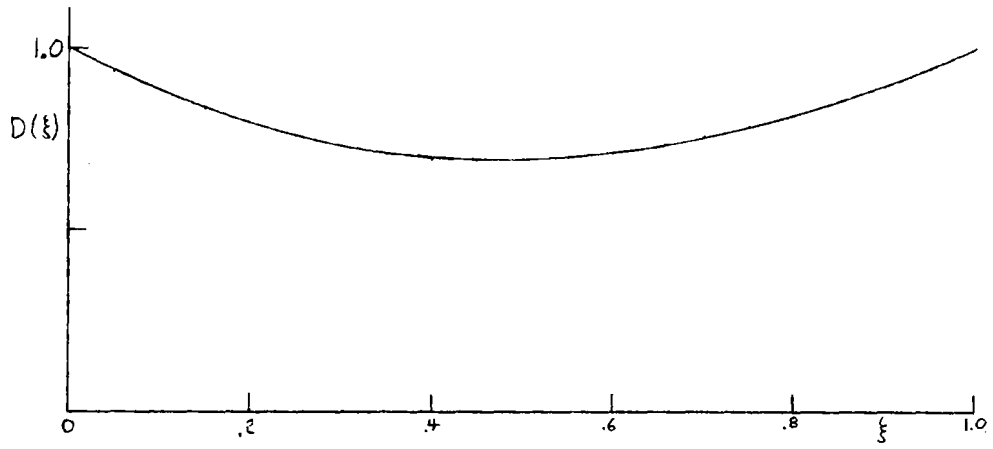
over $(0, \infty)$, we cast the result into the desensitized form

$$(3) \qquad \xi = \frac{1}{1 + .3606032x}$$

$$\phi^*(x) = \frac{.3671626\,\xi - .2272232\,\xi^2 + .8601996\,\xi^3}{1 - 1.3562710\xi + 1.6148087\,\xi^2 - .2585377\,\xi^3}$$

in which a D(1) = 1 condition has been imposed on the second form by proper choice of parameter in the first form.

It is easy to see that (3) is an extremely tractable approximation—what we shall call a 'computable' approximation. The variable 'ξ' is limited to the interval (0, 1) of variation, all coefficients and individual terms in the expression are of reasonable magnitude, and the denominator of the second form behaves very well. This may be seen in the graph below.



3. The Problem of Sensitivity: (Cont'd.)

In very many instances, the least sensitive representation that may be obtained is obtained thru imposing an $D(a) = D(b) = 1$ condition on the denominator when an approximation is made over (a,b).

4. The Problem of Best Fit:

Our criterion of best fit will be that of minimum deviation. Can we then say that our sinusoidal error curves give evidence, in each instance, that no better fit can be obtained? We now give an interesting result for polynomial forms. Our demonstration is a by-product of study concerning the subject of weighting functions. The result is, undoubtedly, a familiar one to workers in the field of polynomial approximation.

Consider the polynomial form

$$(1) \qquad y = a_1 x^{p_1} + a_2 x^{p_2} + \ldots + a_n x^{p_n},$$

with 'n' free parameters 'a_i', and specified powers 'p_i' that are distinct integers $0 \leq p_1 < p_2 < \ldots < p_n$, and 'n+1' points (x_i, y_i), $i = 0(1)n$, for which $0 < x_0 < x_1 < \ldots < x_n$. We introduce the residual notation

$$(2) \qquad T_i = a_1 x_i^{p_1} + \ldots + a_n x_i^{p_n} - y_i,$$

and let 'A_i' denote the value of the determinant obtained from the n x n + 1 matrix

$$(3) \qquad \begin{bmatrix} x_0^{p_1}, & x_1^{p_1}, & \ldots, & x_n^{p_1} \\ x_0^{p_2}, & x_1^{p_2}, & \ldots, & x_n^{p_2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_0^{p_n}, & x_1^{p_n}, & \ldots, & x_n^{p_n} \end{bmatrix}$$

by deletion of the 'x_i' column. A brief argument will show that all 'A_i' are positive.

A proof is by induction on 'n'. Expand any 'A_i' by elements of the last column, and refer to Descartes' rule of signs. There are at most 'n-1' roots to the polynomial in the last column variable, and these are each of the remaining column variables. Our polynomial may then be factored in such a fashion that all factors are easily seen to be positive, and the result follows.

Now, we notice that minimization of

$$(4) \qquad \sum_{i=0}^{n} A_i T_i^2,$$

with respect to the 'a_i', yields the equations of condition

$$(5) \qquad T_i = (-1)^i T_0.$$

Let us denote the residuals in this particular case by the notation 'T_i^*', and

4. The Problem of Best Fit: (Cont'd.)

write

$$T^* = |T_i^*| \tag{6}$$

Let T_i^{**} stand for the residuals that result from any other choice of parameters 'a_i', and write

$$M = \operatorname{Max}_i \left| T_i^{**} \right| . \tag{7}$$

It then follows that

$$T^{*2} \sum_{i=o}^{n} A_i = \sum_{i=o}^{n} A_i T_i^{*2} < \sum_{i=o}^{n} A_i T_i^{**2} \leq M^2 \sum_{i=o}^{n} A_i \tag{8}$$

which says that $T^* < M$. The result may be stated as a theorem.

* * *

Theorem: The minimum deviation solution is such that all deviations are equated in absolute value, and the signs of the deviations alternate, unless the form passed exactly thru the 'n+1' points in which case all deviations are zero.

* * *

This result tells us, in the case of polynomial approximation, that if we obtain an error curve with 'n+1' extremals that obey the conditions of the above theorem, then the approximation cannot be improved. For no other instance of the same form can approximate those particular 'n+1' points of the curve to be approximated as closely as the instance in question.

The reference below has come to our attention, and seems quite pertinent.

Reference:

(1) Th. Motzkin, 'Approximation by Curves of a Unisolvent Family' Bulletin of the American Mathematics Society, Vol. 55, No. 8, Pgs. 789-793.

5. Iterated Rational Forms:

Certain limiting cases of rational approximation may also be of practical interest. Thus we may find a form of the type

$$f(ax) = R\{f(x)\}, \tag{1}$$

in which $R(\lambda)$ is a rational function of 'λ', quite adaptable in special instances. Here 'a' stands, in general, for a positive constant sizeably greater than unity. Quite obviously, the problem of fitting such a form is just that of making an usual rational fit.

Thus, consider the function that satisfies

$$H\left(\frac{3}{2}x\right) = 3H^2(x) - 2H^3(x), \tag{2}$$

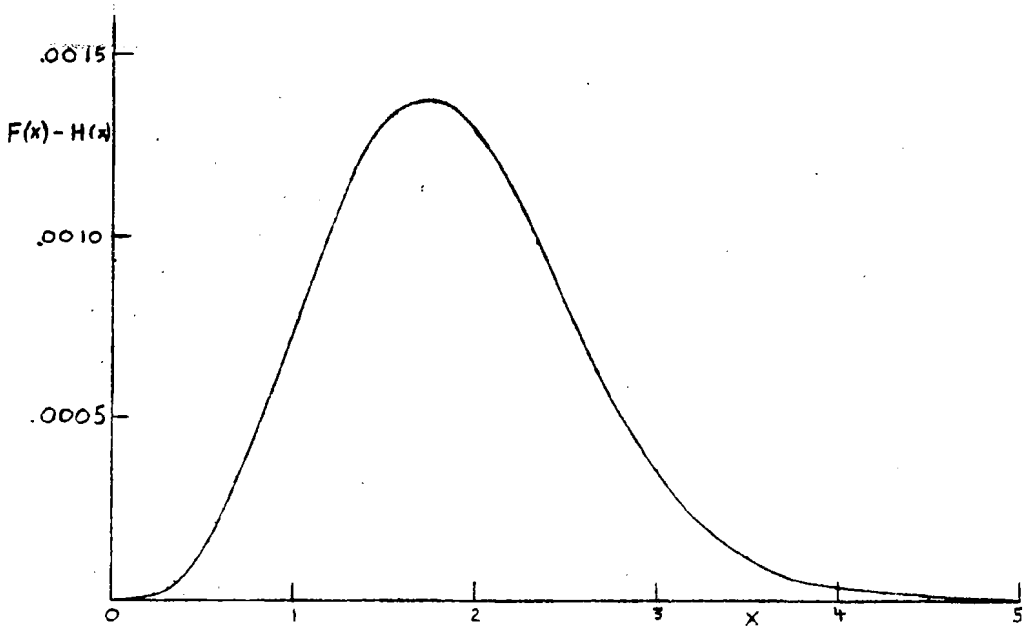
and has the power series expansion

$$H(x) = \frac{1}{2} + \left(\frac{x}{\sqrt{2\pi}}\right) - \frac{16}{15}\left(\frac{x}{\sqrt{2\pi}}\right)^3 + \frac{1024}{975}\left(\frac{x}{\sqrt{2\pi}}\right)^5 + \cdots \tag{3}$$

about the origin. Here $H(0) = \frac{1}{2}$ is a 'fixed point', and $H'(0)$ is an imposed condition. $H(x)$ rather closely approximates the Gaussian error function

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt, \tag{4}$$

over $(-\infty, \infty)$ as may be seen in the difference curve plotted below. Note that $F(x)$ and $H(x)$ both have the same type of odd symmetrical behavior, so that our difference curve need but be given over $(0, \infty)$.



Multivariate Approximations:

We are interested in the subject of a generalized empirical analysis in which multivariate forms of a very general type are studied. These are non-parametric forms in which the unknown quantities are univariate functions of the separate independent variables.

By considering basic functional forms, such as

$$(1) \qquad h^*(x,y) = \frac{f_1(x) + g_1(y)}{1 + f_2(x)g_2(y)},$$

we decompose the problem of multivariate approximation into it's fundamentals. Here $f_1(x)$ and $g_1(y)$ are unknown univariate functions to be determined so that $h^*(x,y)$ shall approximate a specified function $h(x,y)$ over a region of interest as well as possible.

Thus, if the form (1) is inadequate, in an instance, no parametric form of similar structure can be adequate in the same instance. Much work has been done here at RAND on the problem of determining required key values of the unknown functions in such forms. Once tables of values are available for each unknown function, the art of univariate approximation may be employed to complete the fitting process.

A further example of a general functional form is the slide rule form

$$(2) \qquad h(z^*) = f(x) + g(y)$$

in which $z^* = z^*(x,y)$ is to approximate $z(x,y)$ over a specified region. Thus, in a large machine computation that we undertook for a Division of RAND, we approximated

$$(3) \qquad z(x,\gamma) = \sqrt{\left(\frac{2}{\gamma-1}\right)\left(1 - x^{\frac{\gamma-1}{\gamma}}\right)}$$

to 3D over $1.2 \leq \gamma \leq 1.4$, $.05 \leq x \leq 1$ by such a form, and prepared turning point tables of $f(x)$, $g(\gamma)$, and $h(z)$. These were used with efficient success.

Of great practical interest, in the approximation of well behaved bivariate tables, is the simple form

$$(4) \quad z^*(x,y) = f_1(x)\, g_1(y) + f_2(x)\, g_2(y) + \ldots + f_n(x)\, g_n(y)$$

To give an impressive example, for the reader to learn the power of such a form in the fitting of useful tables, we note that

$$(5) \qquad V(h,q) = \frac{1}{2\pi} \int_0^h \int_0^{\frac{qx}{h}} e^{-\frac{1}{2}(x^2+y^2)}\, dy\, dx,$$

6. Multivariate Approximations: (Cont'd.)

may be approximated by

$$(6) \qquad V^*(h,q) = f_1(h)\, g_1\left(\frac{q}{h}\right) + f_2(h)\, g_2\left(\frac{q}{h}\right)$$

over $0 \leq h \leq \infty$, $0 \leq q/h \leq 1$ to 4D. This suffices, with the aid of a table of the error integral, to yield all values of $V(h,q)$.

Reference

(1) H. H. Germond: Miscellaneous Probability Tables, AMP Note No. 14 (Unclassified.)