APPROXIMATION OF PEARSON TYPE IV PROBABILITY INTEGRAL

by

WAYNE ANTHONY WOODWARD, B.A.

A THESIS

IN

MATHEMATICS

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

MASTER OF SCIENCE

May, 1971

ACKNOWLEDGMENTS

I wish to express my appreciation to Professor Thomas L. Boullion for his direction of this thesis and to Professor H. L. Gray for his valuable assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Background

Karl Pearson has derived a family of curves which over the years has proved to be useful in representing practical distributions. These curves are determined by solutions to the differential equation

$$\frac{dg}{dx} = \frac{(x - a)g}{b_0 + b_1 x + b_2 x^2} \qquad (1)$$

In all, Pearson distinguished twelve types or forms of frequency curves which arise as solutions to (1). The solution which will be considered in this thesis is the Pearson Type IV curve. The probability density function corresponding to this curve is given by

$$g(x) = c\left(1 + \frac{x^2}{a^2}\right)^{-m} e^{-k \arctan \frac{x}{a}} \quad -\infty < x < \infty. \qquad (2)$$

The parameters are a, m, and k while c is the constant necessary for the integral over the entire range to equal one. Due to the impossibility of expressing the probability integral in closed form, this curve has been very difficult to use. The purpose of this thesis is to find a relatively simple approximating function for integrals of the form

$$\int_b^\infty g(x)dx$$

where g(x) is given in (2).

Before undertaking this task, it seems to be beneficial to examine further the probability density function itself. The Pearson Type IV distribution can be determined uniquely by its first four moments. The moments about the mean are given by

$$\mu_1' = -\frac{ak}{r}$$

$$\mu_2 = \frac{a^2(r^2 + k^2)}{r^2(r - 1)}$$

$$\mu_3 = -\frac{4a^3k(r^2 + k^2)}{r^3(r - 1)(r - 2)}$$

$$\mu_4 = \frac{3a^4(r^2 + k^2)\{(r + 6)(r^2 + k^2) - 8k^2\}}{r^4(r - 1)(r - 2)(r - 3)}$$

where $r = 2m - 2$. One should note that in order that $\mu_2 > 0$ we must have $r > 1$. Letting $\beta_1 = \mu_3{}^2/\mu_2{}^3$ and $\beta_2 = \mu_4/\mu_2{}^2$, we obtain the following expressions for r, k, and a in terms of the first four moments:

$$r = \frac{6(\beta_2 - \beta_1 - 1)}{(2\beta_2 - 3\beta_1 - 6)}$$

$$k = \frac{r(r - 2)\sqrt{\beta_1}}{\sqrt{\{16(r - 1) - \beta_1(r - 2)^2\}}}$$

$$a = \sqrt{\frac{\mu 2}{16}\{16(r - 1) - \beta_1(r - 2)^2\}}\ .$$

In order to calculate the constant c, it is necessary to evaluate the integral

$$\int_{-\infty}^{\infty} (1 + \frac{x^2}{a^2})^{-m} e^{-k \arctan \frac{x}{a}} dx = \frac{1}{c} .$$

Making the substitutions $\frac{x}{a} = \tan \theta$ and $2m = r + 2$, we have

$$\frac{1}{c} = \int_{-\infty}^{\infty} (1 + \frac{x^2}{a^2})^{-m} e^{-k \arctan \frac{x}{a}} dx$$

$$= \int_{-\pi/2}^{\pi/2} (1 + \tan^2\theta)^{-r-2/2} e^{-k\theta} a \sec^2\theta \, d\theta$$

$$= \int_{-\pi/2}^{\pi/2} a \, e^{-k\theta} \sec^{-r} \theta d\theta$$

$$= a \int_{-\pi/2}^{\pi/2} e^{-k} \cos^r \theta d\theta.$$

Pearson has calculated the integral $\int_{-\pi/2}^{\pi/2} e^{-k\theta} \cos^r\theta d\theta$ and has provided a set of tables [8]. In his notation

$$F(r,k) = \int_{-\pi/2}^{\pi/2} e^{-k\theta} \cos^r\theta d\theta.$$

Thus we have $c = \frac{1}{aF(r,k)}$. The integral $F(r,k)$ exists in closed form when r is an integer, but in other cases Pearson's tables are needed.

The effect which the parameters have upon the curve is also of some interest. Figures 1, 2, and 3 illustrate the effect upon the curve brought about by a change in a parameter. It can be seen that when $k = 0$ the curve is symmetric about the mean which is at $x = 0$ in this case. Figure 1 illustrates the fact that the curve is skew to the right when $k < 0$ and skew to the left when $k > 0$. Of course a change in k also brings about a shift in the mean, since inspection of (2) shows that $a > 0$. Figure 2 indicates that if a and k remain fixed an increase in r brings about a decrease in the variance and a shift in the mean. One can see by Figure 3 that for r and k fixed, an increase in a results in an increase in the variance and a shift in the mean. These results can also be observed by inspection of the density function itself.

Previous Estimations of the Integral

As was mentioned before, the probability integral of the Pearson Type IV curve cannot be found directly. This curve is the only one of Pearson's curves for which the probability integral cannot be reduced to known integrals such as Chi-square integrals or Incomplete Gamma-functions. In Pearson's Tables for Statisticians and Biometricians [9] he expresses his regret for the exception. "The series of tables ought to include Tables of the Incomplete G-Function, i.e., the Probability Integral of the Type IV curve, but the age of the present Editor is likely to preclude his superintending any task, which even exceeds in the magnitude of its calculations that of the Incomplete B-Function."

Karl Pearson was 74 years of age when he made this statement, and, as he said, he never did make such a set of tables. Greenwood and Hartley [5] referring to Pearson's above statement added: "Nor is the project likely to commend itself to a contemporary statistician."

Despite these pessimistic remarks, attempts have been made by L. R. Shenton and J. A. Carpenter [10]. Through the use of Mill's Ratio, Shenton has determined a continued fraction which converges to $\int_b^\infty g(x)dx$ when $\frac{b}{a}$ is positive and to $\int_b^\infty g(x)dx - 1$ when $\frac{b}{a}$ is nega-

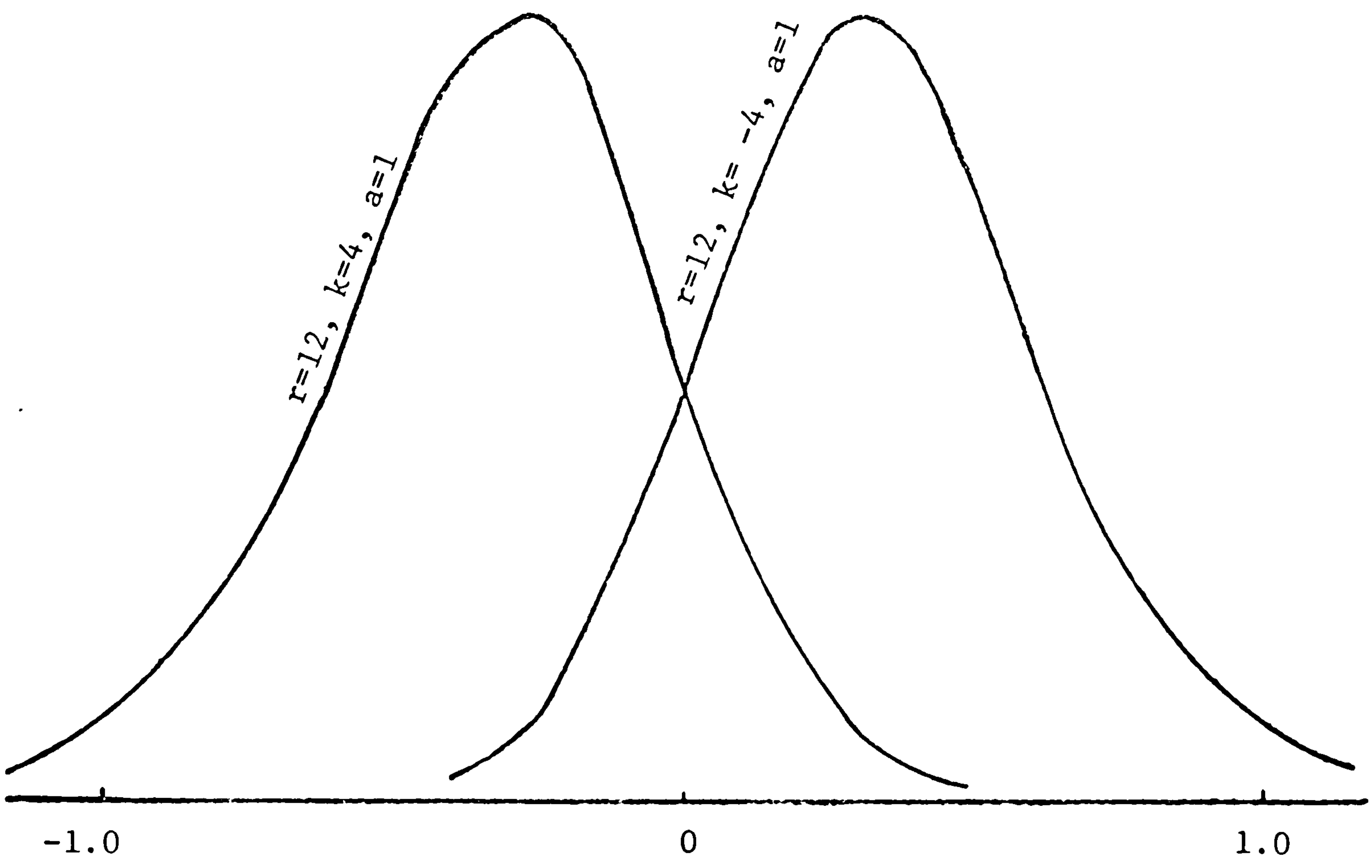

Figure 1. The Effect of a Change in the Parameter k upon the Graph of the Pearson Type IV Probability Curve.

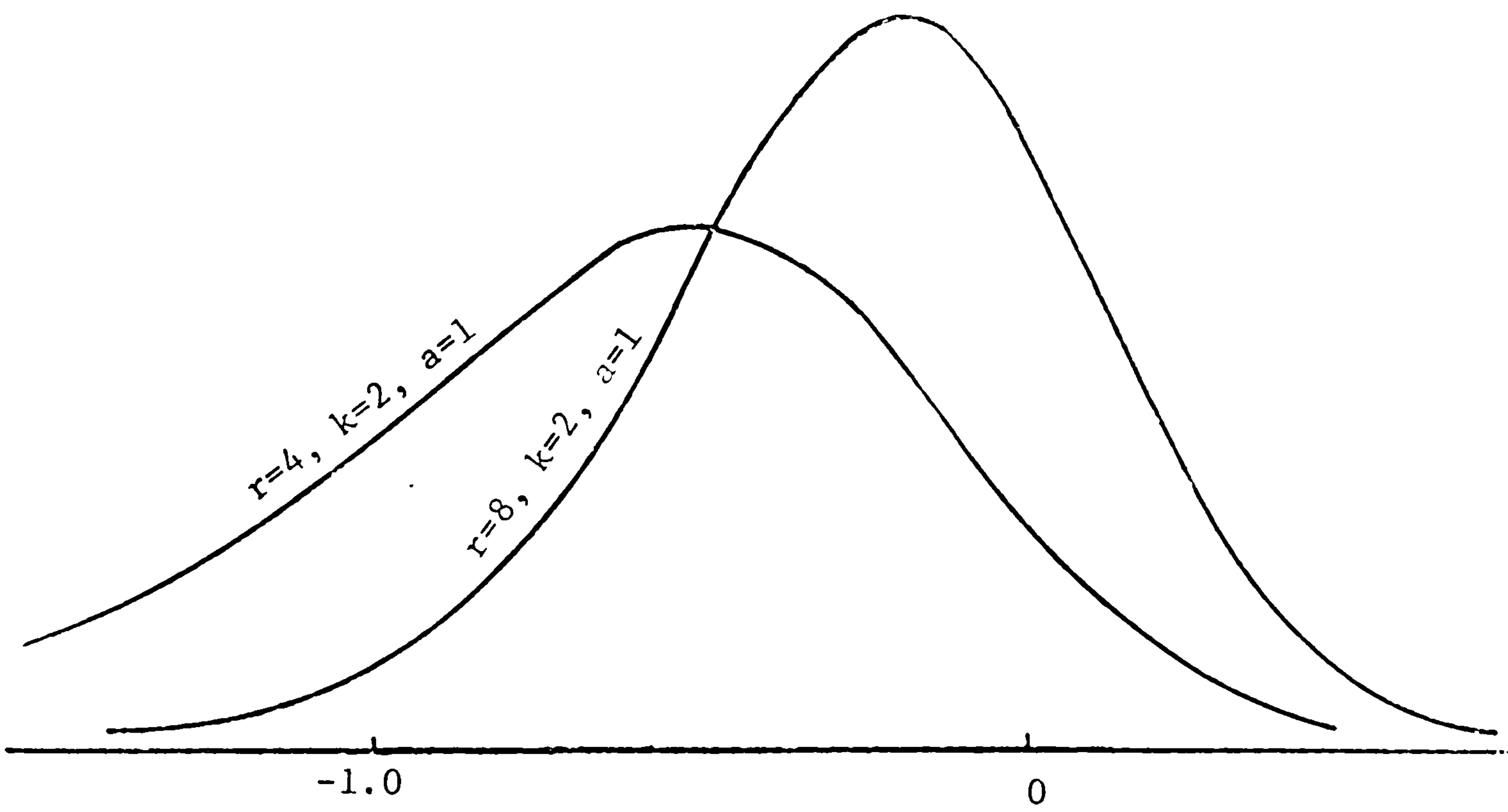

Figure 2. The Effect of a Change in the Parameter r upon the Graph of the Pearson Type IV Probability Curve.

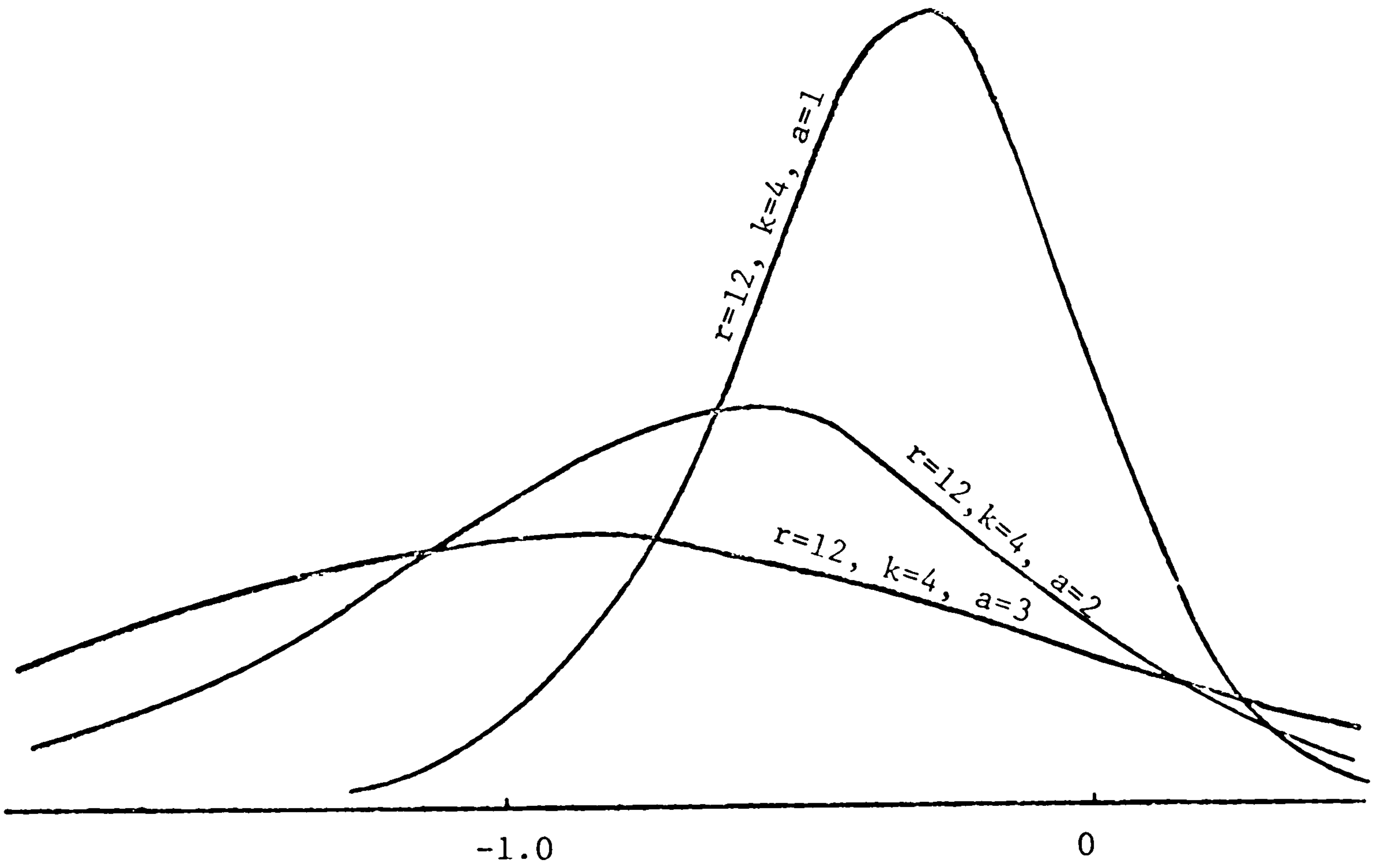

Figure 3. The Effect of a Change in the Parameter a upon the Graph of the Pearson Type IV Probability Curve.

tive. This method, however, does have a few shortcomings. When the value $\frac{b}{a}$ is close to zero, convergence is very slow. This continued fraction expansion, of course, is not a simple approximating function. In the next chapter a truncation of this continued fraction will be made in order to have an approximating function comparable in simplicity to the one proposed here for purposes of comparison.

Another method for evaluating the probability integral of the Pearson Type IV curve has been given by N. L. Johnson and Eric Nixon [6]. These men along with D. E. Amos and E. S. Pearson have published a set of tables of percentage points of Pearson curves. Different types of Pearson curves can be associated with different regions in a graph having $\sqrt{\beta_1}$ and β_2 as the coordinate axes where $\beta_1 = \mu_3^2/\mu_2^3$ and $\beta_2 = \mu_4/\mu_2^2$. Since the shapes of the distributions of the system change continuously across the boundaries of the regions, a single method was used in order to evaluate the percentage points of several types which have adjacent regions on the graph. Among these types is the Pearson Type IV curve. For this group, Johnson and Nixon used the Runge-Kutta-Simpson method to integrate a system of differential equations to evaluate the percentage points of the distributions.

Although this is probably an incomplete listing of results published in this area, it does appear that an extensive set of tables does not exist for the probability integral of the Pearson Type IV curve.

CHAPTER II

DERIVATION OF APPROXIMATING FUNCTIONS

The Use of Nonlinear Transformations in Deriving Approximating Functions

In recent literature several nonlinear transformations have been developed which are useful in evaluating numerically improper integrals of the first kind. In an attempt to obtain an approximating function for the Pearson Type IV probability integral, it seems reasonable to investigate the use of some of these transformations. An attempt will be made to obtain a function which approximates

$$S = \int_b^\infty g(x)dx$$

where g(x) is given in (2). Throughout the remainder of this thesis, S and g(x) will be as above and G(t) will be given by

$$G(t) = \int_b^t g(x)dx.$$

Moreover, f will denote an arbitrary function while F(t) will be given by

$$F(t) = \int_b^\infty f(x)dx,$$

and $F(\infty)$ will be given by

$$F(\infty) = \int_b^\infty f(x)dx.$$

It is desirable for the transformation which will be used to possess certain properties. If A represents a transformation, it is clearly desirable that $A(t) \to S$ as $t \to \infty$. Two additional properties which are desirable for our transformed function to possess are given by the following definition.

Definition 1: If $A(t)$ and $B(t)$ are two sequences of real numbers such that $\lim_{t\to\infty} A(t) = A \neq \pm\infty$ and $\lim_{t\to\infty} B(t) = B \neq \pm\infty$, then we say $A(t)$ converges uniformly better than $B(t)$ on (t_0,∞) if and only if $|A - A(t)| < |B - B(t)|$ for every $t\varepsilon(t_0,\infty)$. Further, if $\lim_{t\to\infty} |\frac{A - A(t)}{B - B(t)}| = 0$, then we say $A(t)$ converges more rapidly than $B(t)$.

Thus, if our transformed function converges uniformly better to S than does $G(t)$, the value of the transformation is closer to S than is $G(t)$ for t sufficiently large. The following theorem shows us the relationship between these two concepts.

Theorem 1: If $A(t)$ converges more rapidly than $B(t)$, then there exists a t_0 such that $A(t)$ converges uniformly better than $B(t)$ on (t_0,∞).

Proof:

By the hypothesis $\lim_{t\to\infty} |\frac{A - A(t)}{B - B(t)}| = 0$. Thus for every $\varepsilon > 0$ there exists a t_ε such that if t is in the interval (t_ε,∞) then $|\frac{A - A(t)}{B - B(t)}| < \varepsilon$. Now, let $\varepsilon = 1$. Then there is a t_0 such that if $t\varepsilon(t_0,\infty)$ then $|\frac{A - A(t)}{B - B(t)}| < 1$ or $|A - A(t)| < |B - B(t)|$.

As can be seen from the proof of Theorem 1, if the transformed function converges more rapidly than G(t), then for a fixed ε, arbitrarily small, there is a t_ε such that if $t \varepsilon (t_\varepsilon, \infty)$ then $|A(t) - S| < \varepsilon \, |G(t) - S|$. Thus it is clear that this would be a desirable property for our transformed function to possess. Also, because of the impossibility of integrating g, the approximating function used must not involve any integration. With these considerations in mind the applicability of certain of these non-linear transformations will be examined.

The H_n Transformations

The first type of transformation to be considered is the H_n transformation which was given by Gray, Atchison, and McWilliams in [3].

Definition 2: Let $f \,\varepsilon\, C^{(2n-1)}$ on (b,∞). Then

$$H_n[F(t)] = \frac{\begin{vmatrix} F & f & f' & \cdots & f^{(n-1)} \\ f & f' & & \cdots & f^{(n)} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ f^{(n-1)} & f^{(n)} & & \cdots & f^{(2n-1)} \end{vmatrix}_t}{\begin{vmatrix} f' & f'' & & \cdots & f^{(n)} \\ f'' & & & & \\ \cdot & & & & \\ \cdot & & & & \\ f^{(n)} & & & \cdots & f^{(2n-1)} \end{vmatrix}_t}$$

where $F(t) = \int_b^t f(x)dx$.

Only $H_1[G(t)]$ and $H_2[G(t)]$ will be considered due to the fact that the higher derivatives involve many terms which would tend to make the approximating function more complicated than desired.

From Definition 2 it follows that $H_1[G(t)]$ is given by

$$H_1[G(t)] = \frac{\begin{vmatrix} G(t) & g(t) \\ g(t) & g'(t) \end{vmatrix}}{\begin{vmatrix} g'(t) \end{vmatrix}}$$

$$= G(t) - \frac{g^2(t)}{g'(t)} .$$

From (2) it follows that

$$g(t) = c(1 + \frac{t^2}{a^2})^{-m} e^{-k \arctan \frac{t}{a}}$$

and

$$g'(t) = c(1 + \frac{t^2}{a^2})^{-m-1} e^{-k \arctan \frac{t}{a}} [- \frac{k}{a} - \frac{2mt}{a^2}].$$

Thus

$$\frac{g^2(t)}{g'(t)} = - \frac{c(1 + \frac{t^2}{a^2})^{-m+1} e^{-k \arctan \frac{t}{a}}}{(\frac{2mt}{a^2} + \frac{k}{a})} ,$$

and

$$H_1[G(t)] = G(t) + \frac{c(1 + \frac{t^2}{a^2})^{-m+1} e^{-k \arctan \frac{t}{a}}}{(\frac{2mt}{a^2} + \frac{k}{a})} .$$

Now $H_1[G(t)]$ will be examined in terms of the considerations mentioned previously. First, Theorem 2 shows that $H_1[G(t)]$ converges to S as $t \to \infty$.

Theorem 2: $H_1[G(t)] \to S$ as $t \to \infty$.

Proof:

We see that $H_1[G(t)] \to S$ as $t \to \infty$ iff $\frac{g^2(t)}{g'(t)} \to 0$ as $t \to \infty$.
But, we have that

$$\frac{g^2(t)}{g'(t)} = \frac{c(1 + \frac{t^2}{a^2})^{-m+1} e^{-k \arctan \frac{t}{a}}}{(- \frac{2mt}{a^2} - \frac{k}{a})}$$

$$= \frac{-ce^{-k \arctan \frac{t}{a}}}{(1 + \frac{t^2}{a^2})^{m-1} (+ \frac{2mt}{a^2} + \frac{k}{a})} .$$

Now, as $t \to \infty$, we see that $-ce^{-k \arctan \frac{t}{a}} \to -ce^{\frac{-k\pi}{2}}$ and $(1 + \frac{t^2}{a^2})^{m-1} (+ \frac{2mt}{a^2}) \to \infty$ since we must have $m > \frac{3}{2}$. Thus $\lim_{t\to\infty} \frac{g^2(t)}{g'(t)} = 0$, and we have that $H_1[G(t)] \to S$ as $t \to \infty$.

Before examining for more rapid convergence, a few observations should be made. We see that $H_1[G(t)]$ is not an acceptable approximating function since it involves G(t). However, if we consider $H_1[G(t)]\big|_{t=b}$, we have $G(b) = 0$, so we eliminate the necessity for integration. However, the question remains as to whether or not $H_1[G(t)]\big|_{t=b}$ is a good approximation. If we know that $H_1[G(t)]$ converges more rapidly to S than does G(t), given an arbitrarily small $\varepsilon > 0$, there is a t_ε such that $|H_1[G(t)] - S| < \varepsilon|G(t) - S|$ for $t\varepsilon(t_\varepsilon,\infty)$. If $b\varepsilon(t_\varepsilon,\infty)$ we would know that $|H_1[G(t)]|_{t=b} - S| < \varepsilon|G(b) - S| = \varepsilon|S|$. Thus we see that if ε is small and b is larger than t_ε, then we should obtain good results with $H_1[G(t)]|_{t=b}$. If $H_1[G(t)]$ does not converge more rapidly than G(t) but does converge uniformly better on an interval (t_0,∞), then we only know that $H_1[G(t)]|_{t=b}$ will be a better estimate for S than G(b) = 0 for b larger than t_0. Even though this does not give us sufficient reason to believe that $H_1[G(t)]|_{t=b}$ will be a good estimate in this case we will use it as an approximating function and compare the results we obtain with those for which we do have more rapid convergence. In both cases we expect to obtain the best results when b is "large", but we do not know how large it must be. In order to assure ourselves that b is "large", in this thesis the approximations will only be applied to integrals such that

$$\int_b^\infty g(t) \le .1.$$

We should first note that $H_1[G(t)]$ does not converge more rapidly than G(t). To accomplish this we need only show that

$$\lim_{t\to\infty} \frac{S - H_1[G(t)]}{S - G(t)} \ne 0.$$

We know that

$$\lim_{t\to\infty} \frac{S - H_1[G(t)]}{S - G(t)} = \lim_{t\to\infty} \frac{S - G(t) + \frac{g^2(t)}{g'(t)}}{S - G(t)}$$

$$= 1 - \lim_{t\to\infty} \frac{g^2(t)}{g'(t)[G(t) - S]} \; .$$

Thus we need to show that

$$\lim_{t\to\infty} \frac{g^2(t)}{g'(t)[G(t) - S]} \neq 1.$$

Using L'Hospital's rule, we have

$$\lim_{t\to\infty} \frac{\frac{g^2(t)}{g'(t)}}{G(t) - S} = \lim_{t\to\infty} \frac{g'(t)2g(t)g'(t) - g^2(t)g''(t)}{[g'(t)]^2 g(t)}$$

$$= \lim_{t\to\infty} \frac{2[g'(t)]^2 - g(t)g''(t)}{[g'(t)]^2}$$

$$= \frac{2m - 1}{2m}$$

$$\neq 1.$$

Even though $H_1[G(t)]$ does not converge more rapidly than $G(t)$, $H_1[G(t)]$ does possess the less restrictive property that it converges uniformly better on an interval (t_0,∞) for some t_0. Gray and Schucany [4] give a necessary and sufficient condition for the existence of such a t_0 to be that $0 < \lim_{t\to\infty} \frac{f^2(t)/f'(t)}{F(t) - F(\infty)} < 2$. From the preceding paragraph we see that

$$\lim_{t\to\infty} \frac{g^2(t)/g'(t)}{G(t) - S} = \frac{2m - 1}{2m},$$

and we know that $0 < \frac{2m - 1}{2m} < 1$. Later in this thesis we will give some transformations which possess the property that the transformed $G(t)$ converges more rapidly than $G(t)$ to S.

Next, we will consider $H_2[G(t)]$. $H_2[G(t)]$ possesses the same properties as does $H_1[G(t)]$ in that it converges to S but does not converge more rapidly than $G(t)$. The criteria for this can be found in [3]. For an approximating function we will use $H_2[G(t)]|_{t=b}$ which is given by

$$H_2[G(t)]|_{t=b} = \frac{\begin{vmatrix} G & g & g' \\ g & g' & g'' \\ g' & g'' & g''' \end{vmatrix}_b}{\begin{vmatrix} g' & g'' \\ g'' & g''' \end{vmatrix}_b}$$

$$= \left. \frac{2gg'g'' - g^2g''' - (g')^3}{g'g''' - (g'')^2} \right|_b .$$

We have

$$g''(x) = c(1 + \frac{x^2}{a^2})^{-m-2} e^{-k \arctan \frac{x}{a}} [(1 + \frac{x^2}{a^2})(\frac{-2m}{a^2}) + \frac{k^2}{a^2}$$

$$+ \frac{2(2m + 1)kx}{a^3} + \frac{4(m + 1)mx^2}{a^4}],$$

and

$$g'''(x) = c(1 + \frac{x^2}{a^2})^{-m-3} e^{-k \arctan \frac{x}{a}} [\frac{k}{a^3}(6m - k^2 + 2) + \frac{6x}{a^4}$$

$$\cdot (m + 1)(2m - k^2) - \frac{6kx^2}{a^5} (2m + 1)(m + 1)$$

$$- \frac{4mx^3}{a^6} (2m + 1)(m + 1)].$$

Thus

$$H_2[G(t)]|_{t=b} = \left. \frac{ce^{-k \arctan \frac{t}{a}} (1 + \frac{t^2}{a^2})^{-m+1} [2AB - C - A^3]}{[AC - B^2]} \right|_b$$

where $A = -\left(\frac{k}{a} + \frac{2mt}{a^2}\right)$

$$B = \frac{k^2 - 2m}{a^2} + \frac{2kt(2m + 1)}{a^3} + \frac{2mt^2(2m + 1)}{a^4}$$

$$C = \frac{k(6m - k^2 + 2)}{a^3} + \frac{6t(m + 1)(2m - k^2)}{a^4}$$

$$- \frac{6kt^2(2m + 1)(m + 1)}{a^5} - \frac{4mt^3(2m + 1)(m + 1)}{a^6}.$$

Some results have been recorded in Table 2. As can be seen, these two approximations give quite good results even though the corresponding transformed functions do not converge more rapidly to S than does G(t).

The L Transformation

Although $H_1[G(t)]|_{t=b}$ and $H_2[G(t)]|_{t=b}$ give satisfactory results, it would seem desirable to use a transformation which does possess the property that the transformed function converges more rapidly than G(t). It is possible to find such a transformation, and we shall follow the line of development of such a transformation below.

We noticed that $H_1[G(t)]$ did not converge more rapidly since $\lim_{t\to\infty} \frac{g^2(t)/g'(t)}{G(t) - S} \neq 1$. Thus it would seem reasonable to examine a transformation of the form $L[F(t)] = F(t) - a\frac{f^2(t)}{f'(t)}$ where a is given by $a = \lim_{t\to\infty} \frac{F(t) - F(\infty)}{f^2(t)/f'(t)}$, and a is assumed to exist. It would seem that the values of the integrals F(t) and F(∞) are necessary in order to evaluate a. However, this is not always true as the following considerations show. If $\lim_{t\to\infty} \frac{f^2(t)}{f'(t)} = 0$, then we may use L'Hospital's rule. Thus we obtain

$$a = \lim_{t\to\infty} \frac{F(t) - F(\infty)}{f^2(t)/f'(t)} = \lim_{t\to\infty} \frac{f(t)}{\frac{2[f'(t)]^2 f(t) - f^2(t)f''(t)}{[f'(t)]^2}}$$

$$= \lim_{t\to\infty} \frac{[f'(t)]^2}{2[f'(t)]^2 - f(t)f''(t)} .$$

One should notice that we may use this alternative form for a since we know that $\lim_{t\to\infty} \frac{g^2(t)}{g'(t)} = 0$. Theorems 3 and 4 show that L[F(t)] possesses the properties which we desire.

<u>Theorem 3:</u> $L[F(t)] \to F(\infty)$ as $t \to \infty$.

<u>Proof:</u>

We see that $L[F(t)] \to F(\infty)$ as $t \to \infty$ iff $a \frac{f^2(t)}{f'(t)} \to 0$ as $t \to \infty$.

We consider two cases:

(1) $a = 0$: The result is obvious.

(2) $a \neq 0$: In this case we have

$$\lim_{t\to\infty} \frac{af^2(t)}{f'(t)} = \lim_{t\to\infty} \frac{F(t) - F(\infty)}{\frac{f^2(t)}{f'(t)}} \frac{f^2(t)}{f'(t)}$$

$$= \lim_{t\to\infty} [F(t) - F(\infty)]$$

$$= 0.$$

<u>Theorem 4:</u> If $a \neq 0$, then $L[F(t)] \to F(\infty)$ more rapidly than $F(t)$ as $t \to \infty$.

<u>Proof:</u>

We need to show that $\lim\limits_{t\to\infty} \frac{F(\infty) - L[F(t)]}{F(\infty) - F(t)} = 0.$

Now, $\frac{F(\infty) - L[F(t)]}{F(\infty) - F(t)} = \frac{F(\infty) - F(t) + a\frac{f^2(t)}{f'(t)}}{F(\infty) - F(t)} = \frac{a\frac{f^2(t)}{f'(t)}}{F(\infty) - F(t)} + 1.$

Thus

$$\lim_{t\to\infty} \frac{F(\infty) - L[F(t)]}{F(\infty) - F(t)} = \lim_{t\to\infty} [1 + \frac{a\frac{f^2(t)}{f'(t)}}{F(\infty) - F(t)}]$$

$$= 1 + a(-\frac{1}{a})$$

$$= 0.$$

It should be noted that more rapid convergence is not achieved if $a = 0$ since in this case $L[F(t)] = F(t)$. Also, one should note that if $a = 1$, then $L[F(t)] = H_1[F(t)]$. The following theorem gives a further criterion for $L[F(t)]$ to reduce to $H_1[F(t)]$.

<u>Theorem 5:</u> If $\lim\limits_{t\to\infty} \frac{f(t)f''(t)}{[f'(t)]^2} = 1$, then $L[F(t)] = H_1[F(t)]$.

Proof:

$$a = \lim_{t\to\infty} \frac{F(t) - F(\infty)}{\frac{f^2(t)}{f'(t)}} = \lim_{t\to\infty} \frac{f(t)}{\frac{f'(t)2f(t)f'(t) - f^2(t)f''(t)}{[f'(t)]^2}}$$

$$= \lim_{t\to\infty} \frac{[f'(t)]^2}{2[f'(t)]^2 - f(t)f''(t)} = \lim_{t\to\infty} \frac{1}{2 - \frac{f(t)f''(t)}{[f'(t)]^2}}$$

$$= 1 \text{ when } \lim_{t\to\infty} \frac{f(t)f''(t)}{[f'(t)]^2} = 1.$$

Now we will consider the use of $L[G(t)]$ as an approximating function. As before, we must use $L[G(t)]\big|_{t=b}$ to avoid having to perform any integration. To evaluate a we notice that
$a = \lim_{t\to\infty} \frac{G(t) - S}{\frac{g^2(t)}{g'(t)}} = \frac{2m}{2m - 1}$ as previously noted.

Thus for the approximating function $L[G(t)]\big|_{t=b}$, we have

$$L[G(t)]_{t=b} = - \frac{2m}{2m - 1} \frac{g^2(t)}{g'(t)}\bigg|_{t=b}$$

$$= \frac{2m}{2m - 1} \frac{c(1 + \frac{t^2}{a^2})^{-m+1} e^{-k \arctan \frac{t}{a}}}{(\frac{2mt}{a^2} + \frac{k}{a})}\Bigg|_{t=b} .$$

The fact that $L[G(t)]$ converges more rapidly than $G(t)$ leads us to expect that beyond a certain point we will get better results with this approximation than with $G(t)$, $H_1[G(t)]$, and $H_2[G(t)]$. However, Table 2 shows us that these results are generally not better in the ranges of b under consideration.

The B_1 Transformation

Another transformation which seems to have possibilities for our use is the B_1 transformation given by Gray and Schucany in [4].

Definition 3: If f is differentiable on (b,∞) and $F(t) = \int_b^t f(x)dx$, then

$$B_1[F(t)] = F(t) - \frac{tf^2(t)}{tf'(t) + f(t)} \quad \text{when } tf'(t) + f(t) \neq 0.$$

This transformation has the properties which we would like our transformation to have as the following theorems show.

Theorem 6: $\lim_{t\to\infty} B_1[G(t)] = S$

Proof:

We need to show that $\dfrac{tg^2(t)}{tg'(t) + g(t)} \to 0$ as $t \to \infty$.

$$\frac{tg^2(t)}{tg'(t) + g(t)}$$

$$= \frac{tc^2(1 + \frac{t^2}{a^2})^{-2m} e^{-2k \arctan \frac{t}{a}}}{tc(1 + \frac{t^2}{a^2})^{-m-1} e^{-k \arctan \frac{t}{a}} (- \frac{k}{a} - \frac{2mt}{a^2}) + c(1 + \frac{t^2}{a^2})^{-m} e^{-k \arctan \frac{t}{a}}}$$

$$= \frac{-ce^{-k \arctan \frac{t}{a}}}{(1 + \frac{t^2}{a^2})^{m-1} (\frac{k}{a} + \frac{2mt}{a^2}) - \frac{(1 + \frac{t^2}{a^2})^m}{t}}.$$

Now as $t \to \infty$, the numerator approaches $-ce - \frac{k\pi}{2}$, and as $t \to \infty$, the denominator approaches ∞. Thus as $t \to \infty$,

$$\frac{tg^2(t)}{tg'(t) + g(t)} \to 0.$$

Now $B_1[G(t)]$ also has the desirable property that $B_1[G(t)]$ converges to S more rapidly than G(t). In order to prove this we must first prove the following lemma.

Lemma 1: If $\lim_{t\to\infty} tf(t) = 0$ and $\lim_{t\to\infty} \frac{f(t)}{tf'(t) + f(t)}$ exists and is finite but not zero, then $B_1[F(t)]$ converges more rapidly than F(t) to $F(\infty)$.

Proof:

We need to show

$$\lim_{t\to\infty} \frac{F(\infty) - B_1[F(t)]}{F(\infty) - F(t)} = \lim_{t\to\infty} \frac{F(\infty) - F(t) + \frac{tf^2(t)}{tf'(t) + f(t)}}{F(\infty) - F(t)}$$

$$= 1 - \lim_{t\to\infty} \frac{tf^2(t)}{[tf'(t) + f(t)][F(t) - F(\infty)]} = 0.$$

Thus we need to show that $\lim_{t\to\infty} \frac{tf^2(t)}{[tf'(t) + f(t)][F(t) - F(\infty)]} = 1.$

$$\lim_{t\to\infty} \frac{tf^2(t)}{[tf'(t) + f(t)][F(t) - F(\infty)]} = \lim_{t\to\infty} \frac{f(t)}{[tf(t)]'} \cdot \frac{tf(t)}{F(t) - F(\infty)}.$$

By hypothesis $\frac{f(t)}{[tf(t)]'}$ exists and is not zero.

Now $tf(t)$ and $F(t) - F(\infty)$ approach zero as $t \to \infty$, so by L'Hospital's rule,

$$\lim_{t\to\infty} \frac{tf(t)}{F(t) - F(\infty)} = \lim_{t\to\infty} \frac{[tf(t)]'}{f(t)} .$$

Thus $\lim_{t\to\infty} \frac{tf^2(t)}{[tf'(t) + f(t)][F(t) - F(\infty)]} = \lim_{t\to\infty} \frac{f(t)}{[tf(t)]'} \cdot \frac{tf(t)}{F(t) - F(\infty)} = 1.$

It should be noted that Lemma 1 is a special case of a result published by Gray and Schucany in [4].

<u>Theorem 7:</u> $B_1[G(t)]$ converges to S more rapidly than G(t).

<u>Proof:</u>

We will show that our function g(t) satisfies the hypothesis of Lemma 1.

First, $\lim_{t\to\infty} tg(t) = \lim_{t\to\infty} ce^{-k \arctan \frac{t}{a}} (1 + \frac{t^2}{a^2})^{-m} \cdot t = 0$ since $m > \frac{1}{2}$ (actually $m > \frac{3}{2}$).

Also

$$\frac{g(t)}{tg'(t) + g(t)} = \frac{c(1 + \frac{t^2}{a^2})^{-m} e^{-k \arctan \frac{t}{a}}}{[tc(1 + \frac{t^2}{a^2})^{-m-1} (- \frac{k}{a} - \frac{2mt}{a^2}) + c(1 + \frac{t^2}{a^2})^{-m}]e^{-k \arctan \frac{t}{a}}}$$

$$= \frac{-1}{t(1 + \frac{t^2}{a^2})^{-1} (\frac{k}{a} + \frac{2mt}{a^2}) - 1} .$$

$$\text{Thus } \lim_{t\to\infty} \frac{g(t)}{tg'(t) + g(t)} = \lim_{t\to\infty} \frac{-1}{t(1 + \frac{t^2}{a^2})^{-1} (\frac{k}{a} + \frac{2mt}{a^2}) - 1}$$

$$= \frac{-1}{2m - 1}$$

$$\neq 0.$$

We see that $\frac{-1}{2m - 1}$ exists since $m \neq \frac{1}{2}$. Thus, since $g(t)$ satisfies the conditions of Lemma 1, we have that $B_1[G(t)]$ converges more rapidly than $G(t)$ to S.

As was done with the preceding transformations, $B_1[G(t)]\big|_{t=b}$ will be used as the corresponding approximating function. Explicitly this function is given by

$$B_1[G(t)]\Big|_{t=b} = \left.\frac{tc(1 + \frac{t^2}{a^2})^{-m} e^{-k \arctan \frac{t}{a}}}{t(1 + \frac{t^2}{a^2})^{-1} (\frac{k}{a} + \frac{2mt}{a^2}) - 1}\right|_{t=b}.$$

It was originally thought that this type of transformation would be our best since it did have the properties desired. However, in the ranges desired, the value of b in the integration was not sufficiently large to give good answers. In fact, as can be seen from Table 2, some of the answers are very poor.

In order to alleviate some of this trouble and still be able to use this transformation, a translation was considered. One should note that $\int_b^\infty f(x)dx = \int_{b+h}^\infty f(x - h)dx = \int_{b+h}^\infty f_1(x)dx$ where

$f_1(x) = f(x - h)$. Now, using our approximating function, we have that

$$-\frac{(b + h)f_1^2(b + h)}{(b + h)f_1'(b + h) + f_1(b + h)}$$ is an approximation for

$\int_b^\infty f(x)dx = \int_{b+h}^\infty f_1(x)dx$. Since $f_1(b + h) = f(b)$, we have that

$$B_h[F(t)] = -\frac{(b + h)f^2(b)}{(b + h)f'(b) + f(b)}$$

should be a good approximating function for $F(\infty)$. The problem next arises as to which values of h give the best results. In order to have one approximating function it was hoped that there would be one value of h which would give good results for all values of the parameters in g(x) and all five of the ranges involved. Empirical results show that such an h does not exist, and even an h which gives good results for all values of the parameters for one range does not exist. Thus it seems that this approximation will not be the one we desire to use.

Continued Fraction Approximations

In order to obtain a relatively simple approximating function with which to compare the results given by those we have derived, we will truncate the continued fraction expansion mentioned in Chapter I. Truncating the expansion after the second convergent gives

$$C_2(t) = \frac{e^{-k \arctan \frac{t}{a}} \cos^{2m-2}(\arctan \frac{t}{a})}{F(r,k)} \frac{2m}{(2m - 1)\ [k + \frac{2mt}{a}]} .$$

Algebraic and trigonometric manipulations reveal that $C_2(t)$ is actually identical to $L[G(t)]$. Thus we see that since convergence of the continued fraction is slow when the value $\frac{t}{a}$ is close to zero, the value of $C_2(t)$ and consequently of $L[G(t)]$ would not be expected to be very accurate in this case. Also, the results should be very poor when $\frac{t}{a}$ is negative since the continued fraction does not converge to the desired value in this case.

For further comparison we will consider the third convergent of the continued fraction. We obtain the expression

$$C_3(t) = \frac{e^{-k \arctan \frac{t}{a}} \cos^{2m-2}(\arctan \frac{t}{a})}{F(r,k)} \quad \frac{4m(m-1)A}{AB + C}$$

where $A = 2(2m+1)[k(m-1) + 2m(m+1)\frac{t}{a}]$,

$$B = 2(2m-1)(m-1)[k + \frac{2mt}{a}],$$

$$C = 4(m-1)(m+1)(2m-1)[k^2 + 4m^2].$$

We see that this expression is comparable in complexity with $H_2[G(t)]$. Table 2 gives a comparison of results given by $C_3(t)$ with those of the other approximations.

CHAPTER III

COMPARISON OF RESULTS

In Chapter II several functions were proposed as candidates for use as approximations for the area under the right tail of a Pearson Type IV distribution. In the present chapter we will present a complete listing of the approximating functions in Table 1 along with a comparison of the accuracy of these functions in Table 2. Finally, based upon the results obtained, recommendations concerning the use of the approximations will be made.

Table 2 gives a comparison of the results given by the approximating functions with a "true value" which is taken to be the convergent of the continued fraction given by Shenton and Carpenter which differs from its predecessor by at most 10^{-6}. Since the convergents eventually straddle the actual true value, we see that our "true value" should be within 10^{-6} of the actual true value.

One should note that the value of the parameter a is always taken to be 1 in Table 2. To illustrate the reasoning for this, we will let X be a random variable with the corresponding density function

$$g_X(x) = \frac{1}{a_o F(r_o, k_o)} \left(1 + \frac{x^2}{a_o^2}\right)^{-m_o} e^{-k_o \arctan \frac{x}{a_o}}$$

which is the Pearson Type IV function with parameters a_o, k_o, and m_o. If we make the change of variable $Y = \frac{X}{a}$, we obtain a random variable Y which has the density function

TABLE 1

APPROXIMATING FUNCTIONS FOR $\int_b^\infty g(x)dx$

TYPE	APPROXIMATION
H_1	$\dfrac{c(1 + \frac{b^2}{a^2})^{-m+1} e^{-k \arctan \frac{b}{a}}}{(\frac{2mb}{a^2} + \frac{k}{a})}$
H_2	$\dfrac{ce^{-k \arctan \frac{b}{a}} (1 + \frac{b^2}{a^2})^{-m+1} [2AB - C - A^3]}{[AC - B^2]}$ *
L	$c(\frac{2m}{2m - 1}) \dfrac{(1 + \frac{b^2}{a^2})^{-m+1} e^{-k \arctan \frac{b}{a}}}{(\frac{2mb}{a^2} + \frac{k}{a})}$
B_1	$\dfrac{cb(1 + \frac{b^2}{a^2})^{-m} e^{-k \arctan \frac{b}{a}}}{b(1 + \frac{b^2}{a^2})^{-1} (\frac{k}{a} + \frac{2mb}{a^2}) - 1}$
C_3	$\dfrac{e^{-k \arctan \frac{b}{a}} \cos^{2m-2} (\arctan \frac{b}{a}) [4m(m - 1)D]}{F(r,k)[DE + F]}$ **

*Footnotes are on the following page.

**Footnotes are on the following page.

$$* \quad A = -\left(\frac{k}{a} + \frac{2mb}{a^2}\right)$$

$$B = \frac{k^2 - 2m}{a^2} + \frac{2kb(2m + 1)}{a^3} + \frac{2mb^2(2m + 1)}{a^4}$$

$$C = \frac{k(6m - k^2 + 2)}{a^3} + \frac{6b(m + 1)(2m - k^2)}{a^4}$$

$$- \frac{6kb^2(2m + 1)(m + 1)}{a^5} - \frac{4mb^3(2m + 1)(m + 1)}{a^6}$$

$$** \quad D = 2(2m + 1)\left[(m - 1)k + 2(m + 1)\frac{mb}{a}\right]$$

$$E = 2(2m - 1)(m - 1)\left[k + \frac{2mb}{a}\right]$$

$$F = 4(m - 1)(2m - 1)(m + 1)\left[k^2 + 4m^2\right]$$

TABLE 2

VALUES OF THE APPROXIMATING FUNCTIONS

Parameters a	r*	k	H_1	H_2	C_3	L	B_1	True Value	$\frac{b}{a}$
1	12	4	.113620	.078528	.071427	.122360	-.060350	.087553	.07
1	8	-4	.106547	.097962	.097095	.118386	.156136	.100113	1.03
1	4	-2	.095937	.094930	.099260	.115125	.152944	.100993	1.25
1	4	2	.121470	.091947	.077568	.145764	-.147605	.099795	.18
1	12	-4	.121716	.102210	.099269	.131079	.184840	.105268	.73
1	8	5.5060	.142924	.078852	.060981	.158804	.055373	.100151	-.17
1	5	0	.108626	.096507	.093037	.126731	.245067	.100093	.59
1	12	4	.058887	.046772	.043898	.063416	-.762461	.049286	.15
1	8	-4	.050222	.049055	.049428	.055803	.066399	.050102	1.24
1	4	-2	.045378	.041595	.050076	.054454	.064162	.050398	1.60
1	4	2	.054486	.048978	.045573	.065383	.221736	.050747	.36
1	12	-4	.053791	.049618	.048950	.057929	.071080	.050259	.88
1	8	5.5060	.062422	.045657	.039628	.069358	.012649	. . .**	-.05
1	5	0	.050433	.048932	.049020	.058839	.080276	.050563	.79
1	12	4	.010837	.009999	.009741	.011671	.016790	.010110	.35
1	8	-4	.009916	.012497	.010444	.011018	.012010	.010477	1.72
1	4	-2	.009133	.010334	.010626	.010960	.011756	.010637	2.5
1	4	2	.010043	.010031	.010391	.012052	.014781	.010644	.75
1	12	-4	.010243	.010151	.010190	.011031	.012209	.010265	1.19
1	8	5.5060	.010933	.009880	.009356	.012147	.053460	.010500	.18
1	5	0	.009403	.011033	.010187	.010971	.012224	.010252	1.28

TABLE 2 - (continued)

Parameters			H_1	H_2	C_3	L	B_1	True Value	$\frac{b}{a}$
a	r	k							
1	12	4	.005231	.004970	.004881	.005633	.007175	.005005	.44
1	8	-4	.004813	.005180	.005140	.005348	.005723	.005150	1.95
1	4	-2	.004387	.004972	.005156	.005264	.005541	.005159	3.00
1	4	2	.004532	.005493	.004896	.005439	.006125	.004960	.95
1	12	-4	.005009	.005009	.005074	.005394	.005848	.005098	1.32
1	8	5.5060	.005255	.004936	.004753	.005839	.010319	.004987	.27
1	5	0	.004689	.005186	.005183	.005470	.005908	.005201	1.50
1	12	4	.001030	.001016	.001010	.001109	.001255	.001022	.61
1	8	-4	.000994	.001070	.001080	.001104	.001153	.001080	2.50
1	4	-2	.000845	.000961	.001004	.001014	.001042	.001004	4.40
1	4	2	.001008	.001124	.001142	.001210	.001268	.001146	1.38
1	12	-4	.000965	.001011	.001002	.001040	.001095	.001004	1.64
1	8	5.5060	.001208	.001188	.001169	.001342	.001664	.001198	.44
1	5	0	.000937	.001040	.001063	.001093	.001136	.001064	2.10

* r = 2m - 2.

**Value unknown due to extremely slow rate of convergence of the continued fraction.
It is approximately .05.

$$g_Y(y) = \frac{1}{a_o F(r_o,k_o)} (1 + y^2)^{-m_o} e^{-k_o \arctan y} \cdot a_o$$

$$= \frac{1}{F(r_o,k_o)} (1 + y^2)^{-m_o} e^{-k_o \arctan y}$$

which is the Pearson Type IV density function with parameters 1, k_o, and m_o for which the table is applicable.

As one can see from Table 2, all of the approximations give the best results at the .01, .005, and .001 levels, which we would expect. The approximation $B_1[G(t)]|_{t=b}$ seems to be the poorest approximation even though $B_1[G(t)]$ possesses the property of more rapid convergence. As can be seen, the results with this approximation do improve for the larger values of b.

Before considering the use of the remaining four approximations, one should check the value of $\frac{b}{a}$. If this value is negative, then $C_3(t)|_{t=b}$ and $L[G(t)]|_{t=b}$ will most likely give very poor results since the continued fraction does not converge to the value required. Actually in this case the values of $H_1[G(t)]|_{t=b}$ and $H_2[G(t)]|_{t=b}$ are also poor, but as can be seen from Table 2 they are better than $C_3(t)|_{t=b}$ and $L[G(t)]|_{t=b}$.

For small positive values of $\frac{b}{a}$, less than .2, the results of the approximations are also relatively poor with $H_1[G(t)]|_{t=b}$ and $H_2[G(t)]|_{t=b}$ again giving the better results, with $H_2[G(t)]|_{t=b}$ giving the better results of the two.

Considering values of $\frac{b}{a}$ greater than .2, we will investigate the accuracy of the approximations. At the .1 level $H_2[G(t)]|_{t=b}$ and $C_3(t)|_{t=b}$ give the best results, varying by less than .007 from the true value while $H_1[G(t)]|_{t=b}$ varies by less than .017 and $L[G(t)]|_{t=b}$ by less than .026. At the .05 level $H_1[G(t)]|_{t=b}$ and $C_3[G(t)]|_{t=b}$ give the best results, each varying from the true value by no more than .005. At the .01 level $C_3(t)|_{t=b}$ gives the best results, varying from the true value by less than .0007 while $H_1[G(t)]|_{t=b}$ and $L[G(t)]|_{t=b}$ vary by less than .0015. $C_3(t)|_{t=b}$ also gives the best results at the .005 and .001 levels varying at the .005 level by less than .0002 and at the .001 level by less than .00003 from the true value. Of the two simpler functions, $H_1[G(t)]|_{t=b}$ gives best results at the .005 level, varying by less than .0008 from the true value; but at the .001 level $L[G(t)]|_{t=b}$ gives the best results, varying by less than .0001. Thus it seems that if great accuracy is not required $H_1[G(t)]|_{t=b}$ should be used for levels other than .001, and at this level and below $L[G(t)]|_{t=b}$ will probably give better results. However, if greater accuracy is required, one should probably use $H_2[G(t)]|_{t=b}$ or $C_3(t)|_{t=b}$ at the .1 level and $C_3(t)|_{t=b}$ for the lower levels.