DISTRIBUTION OF THE RATIO OF CORRELATED SUMS OF GAMMA VARIABLES

John A. Flueck, Burt S. Holland, and Ru-Ying Lee,* Temple University

1. Introduction

In this paper we present the exact probability distribution of the ratio of two correlated gamma random variables for a given bivariate gamma structure. Interest in the theory and application of ratios of random variables has been present in the literature for many years. The first two authors [5] have reviewed some of these results.

The probability distribution of the ratio of two independent r.v.'s is generally easy to obtain, cf. Kendall and Stuart [11, p. 265]. It is well known that the distribution of the ratio of two independent normals is Cauchy and of two independent gammas is the Inverted Beta Type II, Kullback [12]. However, the literature on the distribution of the ratio of two correlated r.v.'s is less developed.

Geary [7] was apparently the first to present results for the ratio of two correlated normals, under the restriction of an always positive denominator. Fieller [4] obtained more general results for the same problem. Rietz [15] presented results for the ratio of correlated uniform r.v.'s. In 1937, Cramér [2] showed that the p.d.f. of the ratio $r=U/V$ of two continuous r.v.'s with $P(V>0)=1$, $E(V)<\infty$, is given by an inversion formula which is in terms of the joint characteristic function of U and V. Application of this approach requires, of course, knowledge of the joint characteristic function and the ability to perform the indicated integration. Gurland [9] has generalized this result by presenting an inversion formula for the probability distribution of a ratio of linear combinations of the same random variables. This formula is based on an n-variate characteristic function. In 1952, C. R. Rao [14, p. 207] showed that under very general conditions the standardized ratio of two means is asymptotically normally distributed. More recently, Marsaglia [13] has studied the probability distribution of the ratio of correlated normal r.v.'s, providing 63 computer-drawn plots of p.d.f.'s to illustrate that the distribution may be symmetric or skewed and unimodal or bimodal.

2. A Bivariate Gamma Distribution

Let X, Y, Z be independently distributed gamma r.v.'s with skew parameters a,b,c and a common scale parameter λ', e.g.,

$$f_X(x)=(\lambda' x)^{a-1}\lambda' e^{-\lambda' x}/\Gamma(a),\ x>0,\ a>0,\ \lambda'>0. \quad (1)$$

Following Weldon's approach, David and Fix [3] defined that for $\lambda'=1$, $U=X+Y$ and $V=X+Z$ are bivariate gamma distributed with density

$$g_{U,V}(u,v) = \frac{e^{-u-v}}{\Gamma(a)\Gamma(b)\Gamma(c)}\int_0^{\min\{u,v\}} t^{a-1}(u-t)^{b-1}(v-t)^{c-1}e^{t}dt \quad (2)$$

and correlation coefficient $\rho=a/[(a+b)(a+c)]^{\frac{1}{2}}$.

In this paper we consider the distribution of the ratio of gamma variables

$$r=U/V=(X+Y)/(X+Z)$$

It is clear that the scale parameter λ' does not appear in the p.d.f. of r, hence we may take $\lambda'=1$ in further discussion of the distribution. Flueck and Holland [5, 6] have previously presented results for the moments of r and a first attempt at its distribution.

A disadvantage of the above bivariate gamma formulation (2) is that it does not admit negative values of the correlation coefficient ρ. The authors are also examining other bivariate gamma distributions (see Johnson and Kotz [10]).

The ratio r easily generalizes to the ratio of sums of gamma variates,

$$r^*= \sum_{i=1}^{m} (X_i+Y_i)/ \sum_{j=1}^{n} (X_j+Z_j),$$

where the $\{X_i\}$, $\{Y_i\}$ and $\{Z_j\}$ are mutually independent r.v.'s identically distributed within each set. If we let $r(a,b,c)$ denote the probability distribution of r with numerator skew parameters a and b, and denominator skew parameters a and c, it can be shown, using the regenerative property of the gamma distribution, that

$$r^* \sim \begin{cases} r(na,\ (m-n)a + mb,\ nc), & m>n \\ r(ma,\ mb,\ (n-m)a + nc), & m\le n, \end{cases} \quad (3)$$

i.e., r^* is a member of the same family of probability distributions as r.

An extension of r is

$$r' = \lambda r = \lambda(X+Y)/(X+Z)$$

so that the numerator is gamma distributed with scale parameter λ'/λ rather than λ'.

3. Derivation of the Probability Density of r.

Let $T=X/(X+Z)$. Then the joint density of r, T, and V is

$$f(r,t,v;a,b,c)=\frac{(tv)^{a-1}e^{-tv}}{\Gamma(a)}\cdot\frac{[v(r-t)]^{b-1}e^{-v(r-t)}}{\Gamma(b)}\cdot\frac{[v(1-t)]^{c-1}e^{-v(1-t)}}{\Gamma(c)}\cdot v^2,\ 0<r,\ 0<v,\ 0<t<\min\{1,r\},$$

and the joint density of r and t is

$$f(r,t;a,b,c)=K(a,b,c)\frac{t^{a-1}(r-t)^{b-1}(1-t)^{c-1}}{(1-t+r)^{a+b+c}},\ 0<r,\ 0<t<\min\{1,r\},$$

where

$K(a,b,c)=\Gamma(a+b+c)/\Gamma(a)\Gamma(b)\Gamma(c)$.

Hence the p.d.f. of r is

$$f(r;a,b,c) = \int_0^{\min\{1,r\}} f(r,t;a,b,c)dt, \; r>0. \qquad (4)$$

Evaluation of (4) by elementary methods appears possible only when a,b,c are all positive integers. For example, when a=b=c=1, (i.e., X, Y, Z each exponential r.v.'s),

$$f(r;1,1,1)=[\max\{1,r\}]^{-2} - (1+r)^{-2}, \; r>0.$$

We attempted to obtain (4) in closed form using the inversion formula in Gurland [9]; however, the integral in the inversion formula has proved difficult to evaluate.

Next we considered calculation of (4) by numerical methods [6], including Simpson and Romberg integration. However, the p.d.f. (4) often has vertical asymptotes at r=0 and r=1; these created problems in obtaining adequate numerical accuracy within reasonable economic limits.

The approach finally adopted and discussed below involved rearrangement of the integral in (4) so that it may be expressed as a weighted difference of hypergeometric functions. The hypergeometric functions are defined and convergent except possibly at r=1.

In (4), for r≤1, replace t with rt to yield

$$f(r;a,b,c) = \begin{cases} K(a,b,c) \cdot \displaystyle\int_0^1 \frac{r^{a+b-1}t^{a-1}(1-t)^{b-1}(1-rt)^{c-1}}{(1-rt+r)^{a+b+c}} dt, & r\le 1, \\ K(a,b,c) \cdot \displaystyle\int_0^1 \frac{t^{a-1}(r-t)^{b-1}(1-t)^{c-1}}{(1-t+r)^{a+b+c}} dt, & r>1. \end{cases} \qquad (5)$$

For r≤1 it follows from the identity

$$r = (1-rt+r) - (1-rt)$$

and 3.211, 9.1821 in [8, pp. 287, 1054] that

$$f(r;a,b,c) = K(a,b,c)\, r^{a+b-2} \int_0^1 \frac{t^{a-1}(1-t)^{b-1}(1-rt)^{c-1}}{(1-rt+r)^{a+b+c-1}} dt - \int_0^1 \frac{t^{a-1}(1-t)^{b-1}(1-rt)^{c}}{(1-rt+r)^{a+b+c}} dt$$

$$= K(a,b,c)B(a,b)r^{a+b-2}(1+r)^{-b-c}[(1+r)F(a,1-c,a+b;r^2) - F(a,-c,a+b;r^2)], \qquad (6)$$

where $F(k_1,k_2,k_3;x)$ is the hypergeometric function (see [1]) and

$$B(k_4,k_5)=\Gamma(k_4)\Gamma(k_5)/\Gamma(k_4+k_5), \; k_4, \; k_5 > 0.$$

Note that F may not converge when r=1. Similarly, when r>1 we use the identity

$$1=(1-t+r)-(r-t)$$

to arrive at

$$f(r;a,b,c)=K(a,b,c)B(a,c)r^{b-a-1}(1+r)^{-b-c}[(1+r) F(a,1-b,a+c;r^{-2})-rF(a,-b,a+c;r^{-2})]. \qquad (7)$$

It is interesting to note that if c in (6) or b in (7) is an integer then the series expression of the corresponding hypergeometric function contains only a finite number of terms.

Unless r=0 or r=1, we have found that (6) and (7) are satisfactory for numerically computing f(r;a,b,c). In most cases less than 100 terms of the hypergeometric series are needed to obtain accuracy to five decimal places, the rate of convergence depending on r, a, b, and c. For the two points r=0 and r=1, analytic formulas have been developed for the p.d.f. and its derivatives.

The presentation here is restricted to a presentation of results; derivations of the analytic formulas rely on the Lebesgue Dominated Convergence Theorem and Fatou's Lemma.

The behavior of f(r;a,b,c) at r=0 is principally governed by the value of a+b. We find:

$$\lim_{r\to 0} f(r;a,b,c) = \begin{cases} 0 & \text{if } a+b>1 \\ c & \text{if } a+b=1 \\ \infty & \text{if } a+b<1 \end{cases} \qquad (8)$$

and

$$\lim_{r\to 0} f'(r;a,b,c) = \begin{cases} 0 & \text{if } a+b>2 \\ c(c+1) & \text{if } a+b=2 \\ \infty & \text{if } 1<a+b<2 \\ c(2a-1-c) & \text{if } a+b=1 \\ -\infty & \text{if } a+b<1 \end{cases} \qquad (9)$$

For k>1, the expression for

$$\lim_{r\to 0} f^{(k)}(r;a,b,c)$$

is quite complicated and we note that the limit may conceivably be positive or negative.

On the other hand, the shape of f(r;a,b,c) in the neighborhood of r=1 is primarily governed by b and c. If b+c>1, then f(r;a,b,c) is continuous at r=1 and

$$f(1;a,b,c)=\frac{\Gamma(b+c-1)}{\Gamma(b)\Gamma(c)} \cdot \frac{2a+b+c-1}{2^{b+c}}, \qquad (10)$$

while if $b+c\le 1$ and $a>0$,

$$f(1^-;a,b,c) = f(1^+;a,b,c) = \infty \qquad (11)$$

If $b+c>2$, $f'(r;a,b,c)$ is continuous at $r=1$ and

$$f'(1;a,b,c)=K(a,b,c)B(a,b+c-2)2^{-b-c-1}[(-a+b-c-2)$$

$$+\frac{a(-a+3b-c-4)}{a+b+c-2} + \frac{a(a+1)(2a+3b+c)}{(a+b+c-2)(a+b+c-1)}]. \qquad (12)$$

Note that (12) may be positive or negative. Next, if $0<b+c\le 1$ and $a>0$ we find

$$f'(1^-;a,b,c)=\infty \qquad f'(1^+;a,b,c)=-\infty \qquad (13)$$

The intermediate case $1<b+c\le 2$, $a>0$ breaks down into six subcases as follows:

If $c>1$ and $b<1$,

$$f'(1^-;a,b,c)= -\infty,\ f'(1^+;a,b,c)= -\infty; \qquad (14a)$$

if $c=1$ and $b=1$,

$$f'(1^-;a,b,c)=\tfrac{1}{4}(2a^2-1),\ f'(1^+;a,b,c)=-\tfrac{1}{4}(2a^2+4a+1); \qquad (14b)$$

if $c=1$ and $b<1$,

$$f'(1^-;a,b,c)=\frac{1}{2^{b+2}}(4a^2+4ab+b^2-4a-3b),$$

$$f'(1^+;a,b,c)=-\infty; \qquad (14c)$$

if $c<1$ and $b>1$,

$$f'(1^-;a,b,c)=\infty,\ f'(1^+;a,b,c)=\infty; \qquad (14d)$$

if $c<1$ and $b=1$,

$$f'(1^-;a,b,c)=\infty,\ f'(1^+;a,b,c)=-\frac{1}{2^{c+2}}(3a^2+3ac+c^2+3a+c); \qquad (14e)$$

if $c<1$ and $b<1$,

$$f'(1^-;a,b,c)=\infty,\ f'(1^+;a,b,c)=-\infty. \qquad (14f)$$

For $a=0$, the independent case, the results for $f(r;a,b,c)$ and $f'(r;a,b,c)$ follow from [12], with

$$f(r;0,b,c) = \frac{1}{B(b,c)}\ \frac{r^{b-1}}{(1+r)^{b+c}}.$$

4. Discussion of the Probability Density of r.

It is seen from (8)-(14) that the shape of $f(r;a,b,c)$ is determined by a, b, and c. To catalogue the situation, we present a 36-cell partitioning of the parameter space $\{(a,b,c): a\ge 0,\ b>0,\ c>0\}$. Table 1 summarizes the implications of (8)-(14) for each cell, while Figures 1-4 contain Calcomp computer plots of $f(r;a,b,c)$ for selected values of a, b, c in order to display the various members of the family. Note that in the Table and in the following discussion we abbreviate $f(r;a,b,c)$ to $f(r)$. Also we use the symbol k to represent a non-negative finite generic constant and k^* to represent a positive finite generic constant.

It is clear from the Figures that the graph of $f(r;a,b,c)$ can assume many unusual shapes and that its appearance is sensitive to small changes in the parameters a, b, c. The following are among the special features of $f(r;a,b,c)$:

1. There is a vertical asymptote at $r=0$ ($f(0;a,b,c)=\infty$ and $f'(0;a,b,c)=-\infty$) iff $a+b<1$.

2. If $a+b\le 1$, the distribution may be bimodal with modes at $r=0$ and $r=1$.

3. There is a vertical asymptote at $r=1$ ($f(1^-;a,b,c)=f(1^+;a,b,c)=f'(1^-;a,b,c)=\infty$ and $f'(1^+;a,b,c)=-\infty$) iff $b+c\le 1$.

4. If $b+c\le 2$, $f'(r;a,b,c)$ is discontinous at $r=1$.

5. The slope of the p.d.f. at $r=0$, $f'(0;a,b,c)$, may be extremely senstitive to small changes in $a+b$. For example, it follows from (9) that

$$f'(0;a,b,c) = \begin{array}{ll} -\infty & \text{if } a+b = 0.99 \\ \pm k & \text{if } a+b = 1.00 \\ \infty & \text{if } a+b = 1.01 \end{array}$$

The Figures also suggest that as a, b, c, each increase, the distribution of r approaches normality. Notice that if in (3) we have $m=n$, the ratio of sums r^* is also a ratio of means, and

$$r^* \sim r(na,nb,nc).$$

Applying a general lemma of C. R. Rao [14], it easily follows that

$$\sqrt{n}\ [r(na,nb,nc)-\frac{a+b}{a+c}]\rightarrow N(0,\frac{(a+b)(b+c)}{(a+c)^3}).$$

This implies that the large sample distribution of $r(na,nb,nc)$ approaches normality as n increases and ultimately becomes degenerate at $(a+b)/(a+c)$.

For completeness we also wish to mention that the first two authors [5] have pointed out that $a+c$ is the quantity which determines whether or not moments of r are finite; in particular, it was found that

$E(r)<\infty$ iff $a+c>1$.

and

$Var(r)<\infty$ iff $a+c>2$.

The fact that some moments of r may be infinite is further evidence of the unusual nature of the distribution of r.

5. Conclusion

This paper has presented exact "closed form" distributional results (6), (7) for the ratio of two correlated gamma r.v.'s. These results also allow the statistician to obtain an efficient numerical approximation to the exact distribution with any degree of accuracy.

The results presented above for the ratio of correlated gamma variates suggest that the probability distributions of ratios of random variables, and of small-sample ratios of sums of random variables, often take unusual forms. In addition to the distribution theory presented here, the p.d.f.'s have an important implication for the practicing statistician: calculation of probability statements involving ratios of r.v.'s is often undertaken assuming normality, but as the above graphs have shown, these calculations must be treated with considerable caution. Furthermore, our results indicate that the accuracy of such probability calculations may rely heavily on precise estimation of the parameters of the ratio distribution. We hope to address this and other problems in future papers as part of our continuing study of ratios of random variables.

References

[1] Abramovitz, M. and Stegun, I. (1964). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Washington: U.S. Government Printing Office.

[2] Cramer, H. (1937). Random Variables and Probability Distributions, (Cambridge Tracts in Mathematics, No. 36), Cambridge University Press.

[3] David, F. N., and Fix, E. (1961). "Rank Correlation and Regression in a Nonnormal Surface," Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, 177-197.

[4] Fieller, E. C. (1932). "The Distribution of the Index in a Normal Bivariate Population," Biometrika, 24, 428-440.

[5] Flueck, J. A., and Holland, B. (1973). "Ratio Estimation Problems in Meterological Research," Proceedings, Third Conference on Probability and Statistics in Atmospheric Science, American Meterological Society, Boston, Mass.

[6] Flueck, J. A., and Holland, B. (1974). "The Moments and Distribution of the Ratio of Two Gamma Variables," 1973 Proceedings, Social Statistics Section, American Statistical Association, Washington, D.C.

[7] Geary, R. C. (1930). "The Frequency Distribution of the Quotient of Two Normal Variates," J. Roy. Statist. Soc., 93, 442-446.

[8] Gradshteyn, I. and Ryzhik, I. (1965). Table of Integrals, Series, and Products. New York: Academic Press.

[9] Gurland, J. (1948). "Inversion Formulae for the Distribution of Ratios," Ann. Math. Statist. 19, 228-237.

[10] Johnson, N. L, and Kotz, S. (1972). Distributions in Statistics: Continuous Multivariate Distributions. New York: John Wiley and Sons, Inc.

[11] Kendall, M. G. and Stuart, A. (1963). The Advanced Theory of Statistics, Vol. I., Second Edition. London: Charles Griffin and Co., Ltd.

[12] Kullback, S. (1936). "The Distribution Laws of the Difference and Quotient of Variables Independently Distributed in Pearson Type III Laws," Ann. Math. Statist. 7, 51-53.

[13] Marsaglia, G. (1965). "Ratio of Normal Variables and Ratios of Sums of Uniform Variables," J. Amer. Statist. Assoc. 60, 193-204.

[14] Rao, C. R. (1952). Advanced Statistical Methods in Biometric Research. New York: John Wiley and Sons, Inc.

[15] Rietz, H. L. (1936). "On the Frequency Distribution of Certain Ratios," Ann. Math. Statist. 7, 145-153.

Footnotes

* The authors' names are listed in alphabetical order.

Table 1. Summary of Behavior of $f(r) = f(r;a,b,c)$ for $r=0,1$.

		$a+b>2$	$a+b=2$	$1<a+b\le 2$	$a+b=1$	$a+b<1$
$b+c>2$		$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1)=\pm k$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1)=\pm k$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1)=\pm k$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=k^*$ $f'(1)=\pm k$	$f(0)=\infty$ $f'(0)=-\infty$ $f(1)=k^*$ $f'(1)=\pm k$
$1<b+c\le 2$	$c>1,\ b<1$	$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1^-)=-\infty$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1^-)=-\infty$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1^-)=-\infty$ $f'(1^+)=-\infty$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=k^*$ $f'(1^-)=-\infty$ $f'(1^+)=-\infty$	$f(0)=\infty$ $f'(0)=-\infty$ $f(1)=k^*$ $f'(1^-)=-\infty$ $f'(1^+)=-\infty$
	$c=1,\ b=1$	$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1^-)=k^*$ $f'(1^+)=-k^*$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1^-)=k^*$ $f'(1^+)=-k^*$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1^-)=\pm k$ $f'(1^+)=-k^*$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=k^*$ $f'(1^-)=-k^*$ $f'(1^+)=-k^*$	Impossible
	$c=1,\ b<1$	$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1^-)=k^*$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1^-)=k^*$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1^-)=\pm k$ $f'(1^+)=-\infty$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=k^*$ $f'(1^-)=-k^*$ $f'(1^+)=-\infty$	$f(0)=\infty$ $f'(0)=-\infty$ $f(1)=k^*$ $f'(1^-)=-k^*$ $f'(1^+)=-\infty$
	$c<1,\ b>1$	$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=\infty$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=\infty$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=\infty$	Impossible	Impossible
	$c<1,\ b=1$	$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-k^*$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-k^*$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-k^*$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-k^*$	Impossible
	$c<1,\ b<1$	$f(0)=0$ $f'(0)=0$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=\infty$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=\infty$ $f'(0)=-\infty$ $f(1)=k^*$ $f'(1^-)=\infty$ $f'(1^+)=+\infty$
$b+c\le 1$		$f(0)=0$ $f'(0)=0$ $f(1)=\infty$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=c(c+1)$ $f(1)=\infty$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=0$ $f'(0)=\infty$ $f(1)=\infty$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=c$ $f'(0)=c(2a-1-c)$ $f(1)=\infty$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$	$f(0)=\infty$ $f'(0)=-\infty$ $f(1)=\infty$ $f'(1^-)=\infty$ $f'(1^+)=-\infty$

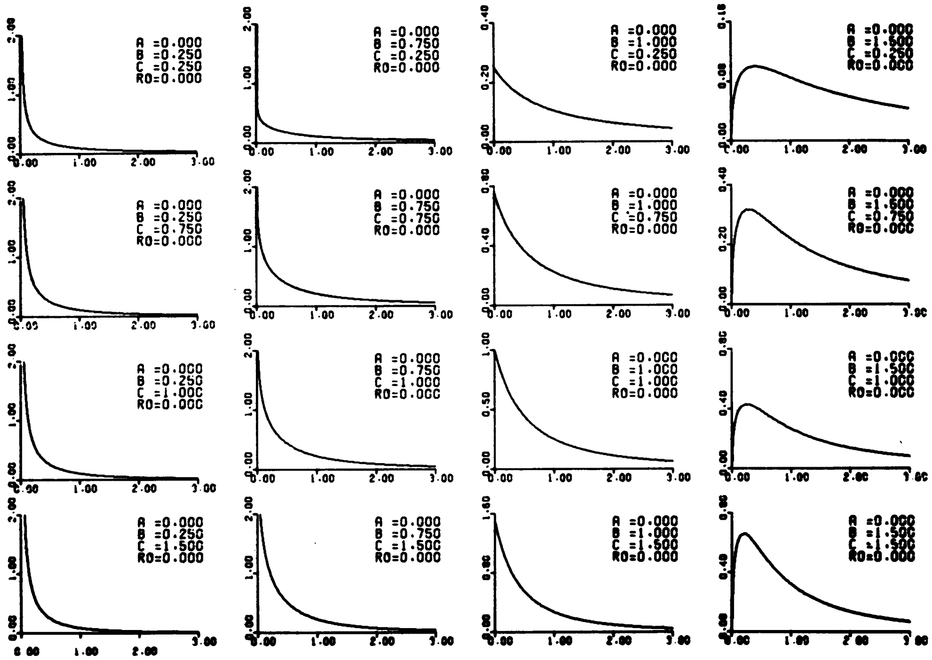

Figure 1. Plots of f(r;a,b,c) versus r.

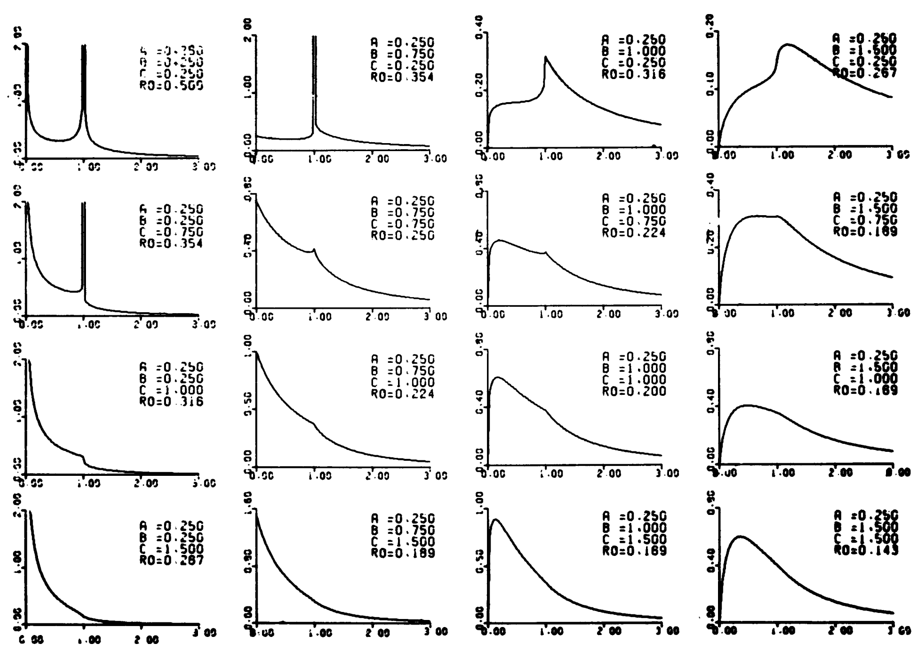

Figure 2. Plots of f(r;a,b,c) versus r.

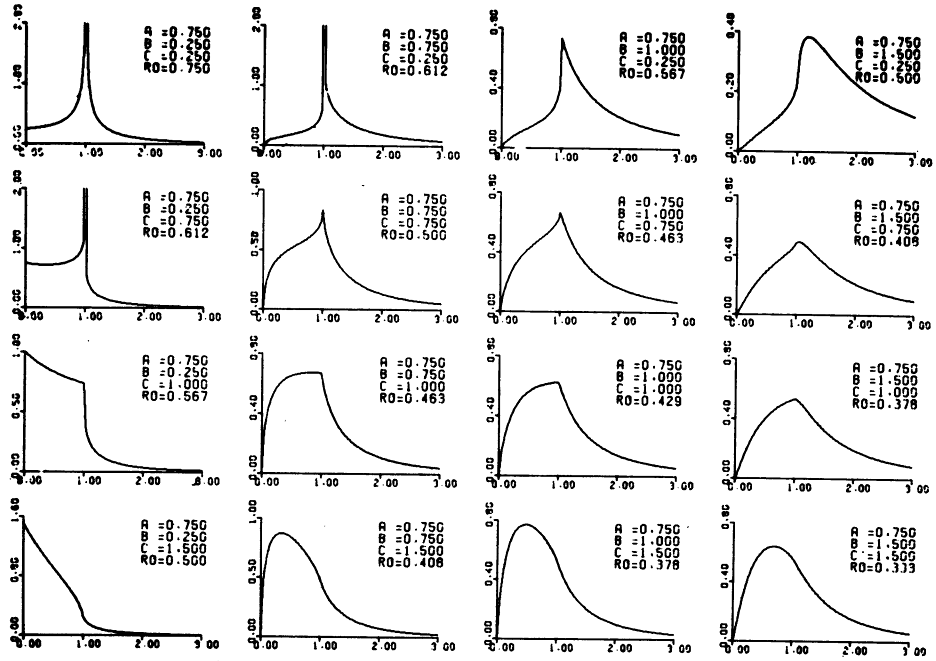

Figure 3. Plots of f(r;a,b,c) versus r.

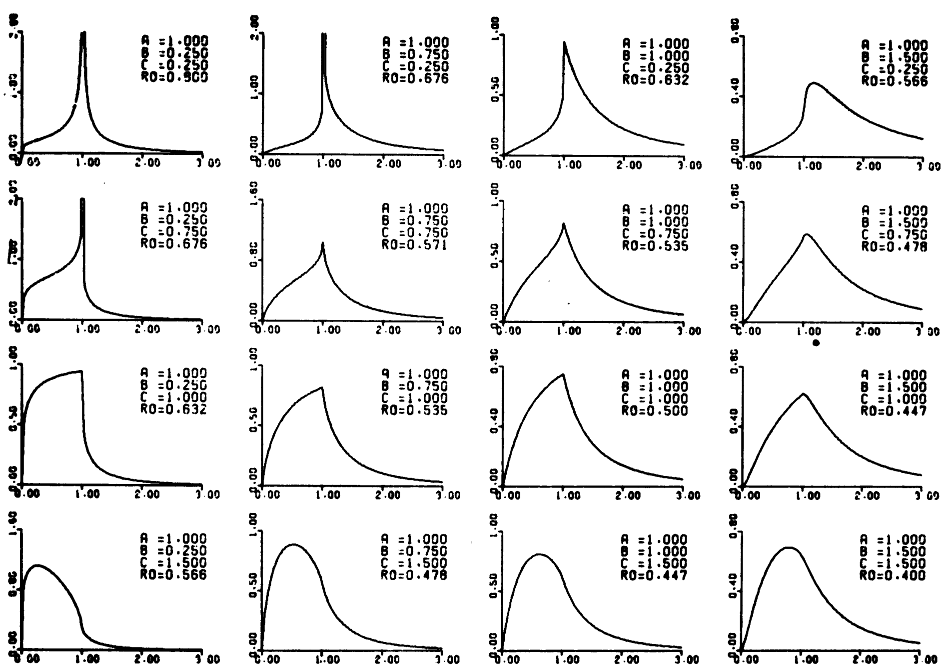

Figure 4. Plots of f(r;a,b,c) versus r.