THE MOMENTS AND DISTRIBUTION OF THE RATIO OF TWO GAMMA VARIABLES

John A. Flueck and Burt S. Holland, Temple University

1. Introduction

We present here work in progress dealing with the ratio of two correlated gamma random variables. Rietz [13] and the authors [4] have presented discussion of some past work in this area.

David and Fix [3] noted that if X, Y, and Z are independent gamma variates with shape parameters a, b, c respectively and common scale parameter λ, e.g.,

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

then the random variables $U = X + Y$ and $W = X + Z$ are bivariate gamma distributed with density

$$f_{U,W}(u,w) = \frac{e^{-u-w}}{\Gamma(a)\Gamma(b)\Gamma(c)}\int_0^{\min\{u,w\}} t^{a-1}(u-t)^{b-1}(w-t)^{c-1} \cdot e^t dt$$

In this paper we shall study the distribution and moments of

$$r' = \frac{X+Y}{X+Z}$$

as an estimator of

$$\frac{E(X+Y)}{E(X+Z)} = \frac{a+b}{a+c}.$$

Since the parameter λ does not appear in the distribution of r' we may henceforth assume without loss of generality that $\lambda = 1$.

The David-Fix formulation of the bivariate gamma distribution constrains the correlation coefficient,

$$\rho = a/[(a+b)(a+c)]^{1/2},$$

to be nonnegative, and thus is somewhat restricted. We are also examining other bivariate gamma distributions (cf. Johnson and Kotz [7]), but since most such distributions are defined in terms of the bivariate characteristic function, the properties of the ratio estimator are more difficult to pursue.

The ratio r' is easily generalized to the ratio of sums of gamma variates,

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

or the ratio of means, $r = nr^*/m$. We assume mutual independence of all r.v.'s in (1). If we let r'(a, b, c) denote the probability distribution of r', it follows that the distribution of r is related to that of r' according to

$$r \sim \begin{cases}(n/m)\cdot r'(na,(m-n)a+mb,nc), & m>n\\ (n/m)\cdot r'(ma,mb,(n-m)a+nc), & m\le n\end{cases} \tag{2}$$

2. Moments of Ratios

The mean and variance of r' are readily computed using well-known properties of the gamma and inverted gamma distribution:

$$E(r') = E[X/(X+Z)] + E(Y)\cdot E(X+Z)^{-1}$$
$$= a(a+c)^{-1} + b(a+c-1)^{-1}, \quad a+c>1.$$

If $a + c \le 1$, E(r') is not finite.

$$E(r')^2 = E[X/(X+Z)]^2 + 2E[X/(X+Z)]\cdot E(X+Z)^{-1} \cdot E(Y) + E(Y^2)E(X+Z)^{-2},$$

hence

$$V(r') = \frac{a(a+1)}{(a+c)(a+c+1)} + 2\frac{a}{(a+c)}\frac{1}{(a+c-1)}b + \frac{b^2+b}{(a+c-1)(a+c-2)} - [E(r')]^2,\ a+c>2.$$

For the variance to be finite it is required that a + c exceed 2.

Using (2), the moments of r are easily obtained from those of r':

$$E(r) = \begin{cases}\frac{a}{a+b} + \frac{nb}{n(a+c)-1}, & n(a+c)>1 \text{ and } m \le n\\[2ex] \frac{n}{m}\left[\frac{a}{a+c} + \frac{(m-n)a+nb}{n(a+c)-1}\right], & n(a+c)>1 \text{ and } m > n\end{cases}$$

$$V(r) = \begin{cases}(n/m)^2\left\{\frac{ma(ma+1)}{n(a+c)[n(a+c)+1]} + \frac{2m^2ab}{n(a+c)[n(a+c)-1]} + \frac{m^2b^2+mb}{[n(a+c)-1][n(a+c)-2]} - \left[\frac{ma}{n(a+c)} + \frac{mb}{n(a+c)-1}\right]^2\right\}, & n(a+c) > 2,\ m\le n.\\[2ex] (n/m)^2\left\{\frac{na(na+1)}{n(a+c)[n(a+c)+1]} + \frac{2a}{(a+c)}\frac{[(n-m)a+mb]}{n(a+c)-1} + \frac{[(n-m)a+mb]^2+[(n-m)a+mb]}{[n(a+c)-1][n(a+c)-2]} - \left[\frac{a}{(a+c)} + \frac{(m-n)a+nb}{n(a+c)-1}\right]^2\right\} & n(a+c)>2,\ m>n\end{cases}$$

It is interesting to note that the expected ratio does not depend upon the numerator sample size if the numerator sample size is not greater than the denominator sample size.

A particular case that has been examined in detail is that of identically distributed numerator and denominator, and equal sample sizes: $a + b = a + c = A$, $\rho = a/A$, $m = n$. Then

$$E(r) = (nA-\rho)/(nA-1), \quad nA>1 \tag{3}$$

and bias of r as an estimator of $(a+b)/(a+c) = 1$ is

$$\text{Bias}\ (r;1) = (1-\rho)/(nA-1), \quad nA>1 \tag{4}$$

Also,

$$V(r) = \left[\frac{\rho(nA\rho+1)}{nA+1} + \frac{2nA\rho(1-\rho)}{nA-1} + \frac{nA(1-\rho)+n^2A^2(1-\rho)^2}{(nA-1)(nA-2)} - \frac{(nA-\rho)^2}{(nA-1)^2}\right], \quad nA>2. \tag{5}$$

Many textbooks dealing with sampling theory, (see, e.g., Cochran [2]), contain the following Taylor Series approximations for the mean and variance of $r = \bar{y}/\bar{x}$:

$$E(r) \doteq R+R\frac{(1-f)}{n^2}\left[\frac{S(X)^2}{\bar{X}^2} - \rho\,\frac{S(X)}{\bar{X}}\,\frac{S(Y)}{\bar{Y}}\right] \tag{6}$$

$$\text{Bias}(r;R) \doteq R\frac{(1-f)}{n^2}\left[\frac{S(X)^2}{\bar{X}^2} - \rho\frac{S(X)}{\bar{X}}\,\frac{S(Y)}{\bar{Y}}\right] \tag{7}$$

$$V(r) \doteq R^2\ \frac{(1-f)}{n^2}\left[\frac{S(X)^2}{\bar{X}^2} + \frac{S(Y)^2}{\bar{Y}^2} - 2\rho\frac{S(X)}{\bar{X}}\,\frac{S(Y)}{\bar{Y}}\right] \tag{8}$$

Here $R = \bar{Y}/\bar{X}$, $\bar{Y}$, and $\bar{X}$ are the population means of Y and X, S(X), and S(Y) are the population standard deviations of X and Y, ρ is the population product moment correlation coefficient between X and Y, and $f = n/N$, the ratio of sample to population size. As is well-known, the expectation bias can be sizable.

If we apply (6-8) to the David-Fix moments (3), (5), assuming $f = 0$, we obtain

$$E(r) \doteq (n^2A - \rho+1)/n^2A \tag{9}$$

$$\text{Bias}\ (r;1) \doteq (1-\rho)/n^2A \tag{10}$$

$$V(r) \doteq 2(1-\rho)/n^2A \tag{11}$$

Comparing the Taylor Series expansion results (6-8) with the exact results (9-11) we see that the conventionally-used Taylor Series Approximations underestimate the expectation bias and variance of r by roughly a multiplicative factor of n.

3. Some Empirical Results

To illustrate the impact of this underestimation, we present examples for two carefully documented weather modification experiments, [5], [14]. Barger and Thom [1] and others have noted that rainfall data are often well-fitted by the gamma distribution. Meteorologists often report the efficacy of cloud seeding experiments via the use of a double ratio, for example,

$$DR = \frac{\text{area one seeded mean}}{\text{area two unseeded mean}} \cdot \frac{\text{area two seeded mean}}{\text{area one unseeded mean}},$$

the product of two independent ratios of means, or its square root (see [9]). Table 1 contains the reported square root of double ratio (RDR); the RDR with each of the ratios corrected for expectation bias using the exact correction (4); and the large sample approximation to the variance of the product of two independent ratios (first reported by J.N.K. Rao[12]) using both the Taylor series approximation for V(r), (8), and the exact value for V(r), (5). It is seen that the RDR corrections are at times appreciable and that the approximate variance of the double ratios employing the exact V(r) is much larger than if the approximate V(r) had been employed.

In reference to their experience based on a considerable number of sampling experiments, and using the Taylor Series approximation for bias and standard error of r, Kish, Namboodiri, and Pillai [8] state, "the ratio of bias of r to standard error of r averaged around .01 and seldom appears greater than .04". In Table 2, we have computed the exact bias to standard error ratio using the exact moments (3), (5) for selected values of ρ, n and A. The selected values are not atypical of rainfall data and may be applicable elsewhere. At least for those parameter values tabulated, the exact bias to standard error ratio is not consistent with the Kish et al statement.

4. Densities

We have recently begun work on the exact probability density of r' (from which the exact probability density of r follows). If we let $V = X/(X+Z)$ and $W = X+Z$, the joint density of r', V and W is easily shown to be

$$g(r',v,w) = \frac{(vw)^{a-1}e^{-vw}}{\Gamma(a)} \cdot \frac{[w(r'-v)]^{b-1}e^{-w(r'-v)}}{\Gamma(b)} \cdot \frac{[w(1-v)]^{c-1}e^{-w(1-v)}}{\Gamma(c)} \cdot w^2,$$

$$0 < w$$
$$0 < r'$$
$$0 < v < \min\{1,r'\}$$

Then the joint density of r' and v is

$$g(r',v) = \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} \cdot \frac{v^{a-1}(r'-v)^{b-1}(1-v)^{c-1}}{(1+r'-v)^{a+b+c}}, \quad \begin{array}{l} 0 < v < \min\{1,r'\} \\ 0 < r' \end{array}$$

and the p.d.f. of r' is

$$g(r') = \int_0^{\min\{1,r'\}} g(r',v)dv, \qquad r' > 0 \qquad (12)$$

In the case where X, Y, and Z are exponential random variables (a=b=c=1), this becomes,

$$g(r') = \begin{cases} 1 - 1/(r'+1)^2, & 0 < r' \le 1 \\ 1/(r')^2 - 1/(r'+1)^2, & r' > 1. \end{cases} \qquad (13)$$

A closed form expression for (12) appears to be readily obtainable only for small integral values of a, b, c. We have also tried to obtain g(r') via an inversion formula in Gurland [6] that uses the joint characteristic function of X, Y and Z, but the integral in the inversion formula has thus far proved elusive.

Instead, a computer program has been developed for the numerical integration needed to generate and plot g(r') and the c.d.f. of r'. Depending on the parameter values under study, a variety of shapes are possible--some sketches of p.d.f.'s appear in Figure 1. These results exhibit both unimodal and bimodal densities with the possibility of an asymptote at one.

The main determinant of the shape of g(r') is the value of b + c. When b + c > 1, the distribution is unimodal and finite for all r' > 0. If $b + c \le 1$,

$$\lim_{r' \to 1} g(r') = \infty.$$

Note that in the distribution of r with $a + b = a + c = A$, $\rho = a/A$, and $m = n$, $b + c$ corresponds to the quantity $2nA(1-\rho)$.

Figure 1(a) displays the p.d.f. in (13), i.e., a = b = c = 1. g(r') has a discontinuous derivative at r' = 1. In Figure 1(b), g(r') has two inflection points below the modal value 1.30. The expected value of both numerator and denominator is larger for the example in Figure 1(c) than in any of the other cases. In the distribution of r this would correspond to a large sample size if A and ρ are small. C.R. Rao [11] has shown that under rather general conditions, the distribution of the ratio of two means is asymptotically normal. This explains the relative lack of shewness in 1(c). In Figure 1(d), the graph of g(r') is monotonically increasing to the left of 1. However, in Figure 1(e), the distribution has a mode between 0 and 1. This might be viewed by some as a bimodal distribution. These results are somewhat similar to the findings of Marsaglia [10] that the distribution of the ratio of two correlated normal random variables (with his assumed bivariate structure) may be bimodal as well as unimodal.

Needless to say, the distribution of r' and probably r is not necessarily well-behaved, and the automatic reliance on the Central Limit Theorem to assure normality for moderate sized samples is highly questionable. Further work is under way.

5. Acknowledgement

The authors wish to acknowledge the assistance of Ru-Ying Lee, Yong Soo Kim, and the Temple University Computer Activity in generating the computer plots.

References

[1] Barger, G.L. and Thom, H.C.S. (1949). "Evaluation of Drought Hazard," Agron. Journ. 41, 519-526.
[2] Cochran, W.G. (1963). Sampling Techniques, Second Edition. New York: John Wiley and Sons, Inc.
[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] Flueck, J. and Holland, B. (1973). "Ratio Estimation Problems in Meteorological Research." Proceedings, Third American Meteorological Society Conference on Probability and Statistics in Atmospheric Science. Boulder, Colo., June 1973.
[5] Godson, W.L., Crozier, C.L. and Holland, J.D. (1966). "An Evaluation of Silver-Iodide Cloud Seeding by Aircraft in West Quebec, Canada, 1961-63," Journ. Appl. Meteorology, 5, 500-512.
[6] Gurland, J. (1948). "Inversion Formulae for the Distribution of Ratios," Annals of Math. Stat., 19, 1948.
[7] Johnson, N.L. and Kotz, S. (1972). Distributions in Statistics: Continuous Multivariate Distributions. New York: John Wiley and Sons, Inc.
[8] Kish, L., Namboodiri, N.K., and Pillai, R.K. (1962). "The Ratio Bias in Surveys," J. Amer. Statist. Assoc. 57, 863-876.
[9] LeCam, L. and Neyman, J. (eds.) (1967). Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. V, Weather Modification. Berkeley: University of California Press.
[10] Marsaglia, G. (1965). "Ratios of Normal Variables and Ratios of Sums of Uniform Variables," J. Amer. Statist. Assoc., 60, 193-204.
[11] Rao, C.R. (1952). Advanced Statistical Methods in Biometric Research. New York: John Wiley and Sons, Inc.
[12] Rao, J.N.K. (1957). "Double Ratio Estimate in Forest Surveys," Journ. Indian Soc. Agr. Stat., 9, 191-204.
[13] Rietz, H.L. (1936). "On the Frequency Distribution of Certain Ratios," Ann. Math. Statist. 7, 145-153.
[14] Smith, E.J. (1967). "Cloud Seeding Experiments in Australia," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, V, 161-76. Univ. of California Press.

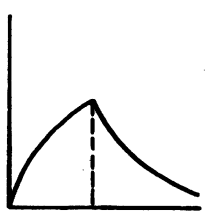
Table 1. Environmental Results from Cloud Seeding Experiments

Project		n	RDR	CRDR	AVDR1	AVDR2
West Quebec	1960-63	23	.9310	.9154	.0022	.0566
	1960	6	.9275	.8560		
	1961	5	.6801	.6182		
	1962	5	1.793	1.695		
	1963	7	.8618	.8127		
South Australia	1957-59	22	.9525	.9473	.0008	.0174
	1957	8	.9711	.9563		
	1958	9	.9234	.9168		
	1959	5	.8836	.8482		

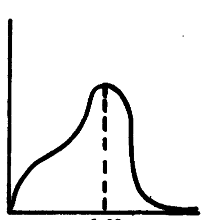
Key:
RDR = square root of product of two independent ratios
CRDR = RDR with each of ratios corrected for expectation bias using $(1 - \rho)/(nA - 1)$
AVDR1 = large sample approximation to variance of product of two independent ratios, [12], inserting $V(r) = 2(1 - \rho)/n^2A$ in the approximation
AVDR2 = same as AVDR1 except using exact value for $V(r)$

Table 2. Relative Bias of r for Selected Values of ρ, n, A

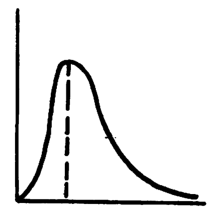
ρ	n	A	bias/σ
.00	10	.25	0.224
.00	10	2.0	0.152
.00	50	.25	0.187
.00	50	2.0	0.0702
.50	10	.25	0.188
.50	10	2.0	0.110
.50	50	.25	0.137
.50	50	2.0	0.0505
.95	10	.25	0.0756
.95	10	2.0	0.0358
.95	50	.25	0.0454
.95	50	2.0	0.0158



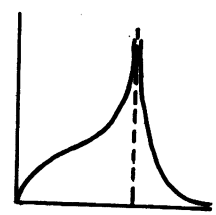
(a) b+c > 1
ρ = 0.50
E(X+Y) = 2.00
E(X+Z) = 2.00



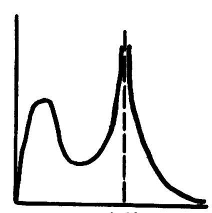
(b) b+c > 1
ρ = 0.50
E(X+Y) = 2.00
E(X+Z) = 1.00



(c) b+c > 1
ρ = 0.50
E(X+Y) = 8.00
E(X+Z) = 8.00



(d) b+c = 1
ρ = 0.75
E(X+Y) = 2.00
E(X+Z) = 2.00



(e) b+c < 1
ρ = 0.75
E(X+Y) = 0.50
E(X+Z) = 0.50

Figure 1. Selected Sketches of Computer Plots of g(r')