NOTES

A GENERALIZATION OF THE GAMMA DISTRIBUTION

By E. W. Stacy

IBM Development Laboratory, Endicott, N. Y.

1. Summary and introduction. This paper concerns a generalization of the gamma distribution, the specific form being suggested by Liouville's extension to Dirichlet's integral formula [3]. In this form it also may be regarded as a special case of a function introduced by L. Amoroso [1] and R. d'Addario [2] in analyzing the distribution of economic income. (Also listed in [4] and [5].)

In essence, the generalization (1) herein is accomplished by supplying a positive parameter, p, as an exponent in the exponential factor of the gamma distribution. The moment generating function is shown, and cumulative probabilities are related directly to the incomplete gamma function (tabulated in [6]).

Distributions are given for various functions of independent "generalized gamma variates" thus defined, special attention being given to the sum of such variates. Convolution results occur in alternating series form, with coefficients whose evaluation may be tedious and lengthy. An upper bound is provided for the modulus of each term, and simplified computation methods are developed for some special cases. A corollary is derived showing that the researches of Robbins in [7] apply to a larger class of problems than was treated in [7]. Extensions of his methods lead to iterative formulae for the coefficients in series obtained for an even larger class of problems.

2. A frequency function and some elementary properties. Let X be a random variable whose frequency function is

$$f(x; a, d, p) = (p/a^d)x^{d-1}e^{-(x/a)^p}/\Gamma(d/p), \tag{1}$$

for non-negative values of x and positive values of the parameters a, d, and p. The familiar gamma, Chi, Chi-squared, exponential, and Weibull variates are special cases, as are certain functions of a standard normal variate—viz., its positive even powers, its modulus, and all positive powers of its modulus.

The function (1) may be regarded as a generalization of the gamma distribution and elementary properties of the variable X may be verified directly. Denoting its cumulative frequency by $F(x; a, d, p)$ and its moment generating function by $M(\theta)$, we have

$$F(x; a, d, p) = \Gamma_z(d/p)/\Gamma(d/p) \tag{2}$$

Received July 31, 1961; revised January 7, 1962.

and

$$M(\theta) = \sum_{r=0}^{\infty} \frac{(\theta a)^r}{r!} \left[\Gamma\left(\frac{d+r}{p}\right) \Big/ \Gamma(d/p) \right], \tag{3}$$

where $z = (x/a)^p$ and

$$\Gamma_z(d/p) = \int_0^z v^{(d/p)-1} e^{-v} \, dv. \tag{4}$$

Also, with $m > 0$, when X is distributed by (1),

$$(\partial/\partial y') \text{ Prob } (X^m \leqq y') = f(y'; a^m, d/m, p/m). \tag{5}$$

3. Distributions for some functions of independent variables. Suppose now that X and Y are independently distributed with respective frequency functions $f(x; a_1, d_1, p)$ and $f(y; a_2, d_2, p)$. We consider the random variable

$$T = X/Y. \tag{6}$$

Straightforward integration techniques indicate that if

$$W = T^p/[T^p + (a_1/a_2)^p], \tag{7}$$

W is a Beta variable with parameters (d_1/p) and (d_2/p). Since W is a strictly increasing function of T, we can deduce cumulative probabilities for T; i.e.,

$$\text{Prob}(T \leqq t) = \text{Prob}(W \leqq \{t^p/[t^p + (a_1/a_2)^p]\}). \tag{8}$$

It is evident also that cumulative probabilities for positive powers of T are available from similar considerations.

Next let $X_1, \cdots, X_n$ be an independent set of random variables, X_i having frequency

$$f(x_i; a_i, d_i, p_i), \qquad i = 1, \cdots, n \tag{9}$$

Further define

$$Z_n = \sum_{i=1}^{n} (X_i/a_i)^{p_i}, \tag{10}$$

with frequency $f(z_n)$. Then

$$f(z_n) = f\left(z_n; 1, \sum_{i=1}^{n} (d_i/p_i), 1\right); \tag{11}$$

and the distribution of any positive power of Z_n is evident from foregoing considerations.

4. Distribution for the sum of independent variables. The remainder of the discussion concerns the sum

$$Y = \sum_{i=1}^{n} X_i, \tag{12}$$

where the X_i are defined as in the previous paragraph. We denote the cumulative distribution function of Y by $G_n(y)$ and the corresponding frequency function by $g_n(y)$. After expanding the joint frequency of $X_1, \cdots, X_n$, direct integration gives

$$G_n(y) = \beta_n y^d \sum_{j=0}^{\infty} (-1)^j A_j, \tag{13}$$

where

$$d = \sum_{i=1}^{n} d_i, \qquad \beta_n = \prod_{i=1}^{n} \left[p_i / a_i^{d_i} \Gamma\left(\frac{d_i}{p_i}\right) \right], \tag{14}$$

and

$$A_j = \sum_{k_1+\cdots+k_n=j} \frac{y^{\sum_{i=1}^{n} p_i k_i}}{\Gamma\left(d + \sum_{i=1}^{n} p_i k_i + 1\right)} \prod_{i=1}^{n} \frac{\Gamma(d_i + p_i k_i)}{k_i!\, a_i^{p_i k_i}}. \tag{15}$$

Differentiation of (13) with respect to y leads to

$$g_n(y) = \beta_n y^{d-1} \sum_{j=0}^{\infty} (-1)^j \bar{A}_j, \tag{16}$$

where

$$\bar{A}_j = \sum_{k_1+\cdots+k_n=j} \frac{y^{\sum_{i=1}^{n} p_i k_i}}{\Gamma\left(d + \sum_{i=1}^{n} p_i k_i\right)} \prod_{i=1}^{n} \frac{\Gamma(d_i + p_i k_i)}{k_i!\, a_i^{p_i k_i}}. \tag{17}$$

No general methods have been found for easy evaluation of the quantities $\bar{A}_j$. Computation of these coefficients, however, can be simplified in certain special cases.

5. Simplified computations—a special case. In what follows, we assume that all the p_i have unit value and denote the result of substituting these values in (16) by $h_n(y)$. Then

$$h_n(y) = y^{d-1} \sum_{j=0}^{\infty} \frac{(-y)^j}{\Gamma(d+j)} \sum_{k_1+\cdots+k_n=j} \prod_{i=1}^{n} \frac{\Gamma(d_i + k_i)}{k_i!\, \Gamma(d_i) a_i^{d_i+k_i}}. \tag{18}$$

When the a_i have a common value a, say, we have the alternative expression

$$\begin{aligned} h_n(y) &= f(y; a, d, 1) \\ &= a^{-d} y^{d-1} e^{-y/a} / \Gamma(d) \\ &= a^{-d} y^{d-1} \sum_{j=0}^{\infty} [(-y/a)^j / (j! \Gamma(d))]. \end{aligned} \tag{19}$$

Replacing the quantities a_i in (18) by their common value, and comparing coefficients of (18) and (19), it is seen that

$$\sum_{k_1+\cdots+k_n=j} \prod_{i=1}^{n} \frac{\Gamma(d_i + k_i)}{k_i!\, \Gamma(d_i)} = \frac{\Gamma(d+j)}{j!\, \Gamma(d)}. \tag{20}$$

The magnitude of the $(j + 1)$st term of (18) consequently does not exceed

$$(y/\alpha)^{d+j}/(j!\Gamma(d)), \tag{21}$$

where $\alpha = \min (a_1, \cdots, a_n)$.

A corollary stemming from (20) is

$$\sum_{t_1+\cdots+t_s=j} \prod_{k=1}^{s} \frac{\Gamma(m_k + t_k)}{t_k\,!\,\Gamma(m_k)} = \sum_{k_1+\cdots+k_n=j} \prod_{i=1}^{n} \frac{\Gamma(d_i + k_i)}{k_i\,!\,\Gamma(d_i)} \tag{22}$$

so long as the m_k are positive and $\sum_{k=1}^{s} m_k = \sum_{i=1}^{n} d_i = d$.

Now let

$$d_i = \nu_i/\gamma_i = \tilde{\nu}_i/\gamma, \tag{23}$$

where ν_i and γ_i are positive integers; γ is the least common multiple of $\gamma_1, \cdots, \gamma_n$; the $\tilde{\nu}_i$ are determined according to the last part of (23); and $i = 1, \cdots, n$. Combining (18) and (22), we have

$$h_n(y) = y^{d-1} \sum_{j=0}^{\infty} \{(-y)^j/[\Gamma(d + j)]\} \bar{\bar{A}}_j, \tag{24}$$

where

$$\bar{\bar{A}}_j = [\Gamma(1/\gamma)]^{-s} \sum_{t_1+\cdots+t_s=j} \prod_{\theta=1}^{s} \frac{\Gamma((1/\gamma) + t_\theta)}{t_\theta\,!\,\tilde{a}_\theta^{(1/\gamma)+t_\theta}} \tag{25}$$

in which

$$s = \sum_{i=1}^{n} \tilde{\nu}_i, \tag{26}$$

$$\tilde{a}_\theta = \begin{cases} a_1, & \theta = 1, \cdots, \tilde{\nu}_1 \\ a_i, & \theta = \sum_{m=1}^{i-1} \tilde{\nu}_m + 1, \cdots, \sum_{m=1}^{i} \tilde{\nu}_m;\ i = 2, \cdots, n. \end{cases} \tag{27}$$

The methods of Robbins [7] facilitate the evaluation of (24) when γ equals 2. It is evident, therefore, that his methods apply to (18) when the values of the d_i are positive integer multiples of $\frac{1}{2}$. We shall develop methods which apply to (24) for arbitrary, positive, integer values of γ. Those methods will apply to (18) when the values of the d_i are positive and rational.

We continue then by letting

$$X_1, \cdots, X_n; \quad Y_1, \cdots, Y_n; \quad \cdots; \quad Z_1, \cdots, Z_n$$

be β independently and identically distributed sets of n random variables each, the individual sets being distributed by

$$\prod_{i=1}^{n} f(s_i; a_i, d_i, 1). \tag{28}$$

Here, the variable s_i is interpreted as x_i when referring to the X_i, as y_i when referring to the Y_i, etc. We specify also that the d_i are rational as defined by (23) and make use of the fact that there is no loss of generality if we set the d_i equal to $1/\gamma$.

Defining

$$X_{\beta,n} = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i + \cdots + \sum_{i=1}^{n} Z_i, \tag{29}$$

$$P_{\beta,n}(x) = \text{Prob}\,[X_{\beta,n} \leqq x],$$

and

$$p_{\beta,n}(x) = (\partial/\partial x) P_{\beta,n}(x), \tag{30}$$

we write from (24)

$$p_{1,n}(x) = \frac{x^{(n/\gamma)-1}}{(a_1 a_2 \cdots a_n)^{1/\gamma}} \sum_{j=0}^{\infty} \frac{(-x)^j}{\Gamma((n/\gamma) + j)} c_j, \tag{31}$$

in which

$$c_j = [\Gamma(1/\gamma)]^{-n} \sum_{k_1+\cdots+k_n=j} \prod_{i=1}^{n} \frac{\Gamma((1/\gamma) + k_i)}{k_i\,!\, a_i^{k_i}}. \tag{32}$$

Convolution techniques using (31) lead to

$$p_{\gamma,n}(x) = \frac{x^{n-1}}{a_1 a_2 \cdots a_n} \sum_{j=0}^{\infty} \left\{ \sum_{k_1+\cdots+k_\gamma=j} \left(\prod_{i=1}^{\gamma} c_{k_i} \right) \right\} \frac{(-x)^j}{\Gamma(n+j)}. \tag{33}$$

Alternatively, setting

$$q_i = \prod_{j \neq i} (a_j - a_i)^{-1} \tag{34}$$

and assuming as a temporary expedient that the a_i are distinct, we write

$$\begin{aligned} p_{\gamma,n}(x) &= (-1)^{n-1} \sum_{i=1}^{n} q_i a_i^{n-2} e^{-x/a_i}, \\ &= (-1)^{n-1} \sum_{j=0}^{\infty} \left\{ \sum_{i=1}^{n} q_i a_i^{n-j-2} \right\} (-x)^j/j!. \end{aligned} \tag{35}$$

Comparison of coefficients in (33) and (35) establishes

$$\sum_{k_1+\cdots+k_\gamma=j} \left(\prod_{i=1}^{\gamma} c_{k_i} \right) = (a_1 \cdots a_n) \sum_{i=1}^{n} q_i a_i^{-(j+1)} \tag{36}$$

for $j = 0, 1, 2, \cdots$.

Now c_0 can be determined by definition (32) and we have $c_0 = 1$

$$\gamma c_1 = (a_1 \cdots a_n) \sum_{i=1}^{n} (q_i/a_i^2),$$

$$\cdots$$

$$\gamma c_j = (a_1 \cdots a_n) \sum_{i=1}^{n} (q_i/a_i^{j+1}) - \sum_{t=2}^{\phi} \binom{\gamma}{t} \sum_{s_1+\cdots+s_t=j} \left(\prod_{r=1}^{t} c_{s_r} \right)$$

in which the s_i are positive with $s_r \geqq s_{r+1}$, $r = 1, \cdots, t - 1$, and $\phi = \min(j, \gamma)$.

Cases in which the a_i are not distinct can be treated as above except that (36) must be replaced by the corresponding limit formulae.

Acknowledgment. The author wishes to thank D. L. Heck for constructive suggestions regarding the preparation of this paper.

REFERENCES

[1] AMOROSO, L. (1925). Ricerche intorno alla curva dei redditi. *Ann. Mat. Pura Appl.* Ser. 4 **21** 123–159.

[2] D'ADDARIO, R. (1932). Intorno alla curva dei redditi di Amoroso. *Riv. Italiana Statist. Econ. Fnanza*, anno 4, No. 1.

[3] EDWARDS, JOSEPH (1954). *A Treatise on the Integral Calculus.* Chelsea Pub., New York.

[4] HAIGHT, FRANK A. (1961). Index to the distributions of mathematical statistics. *J. Research*, National Bureau of Standards, Sec. B **65B** No. 1, 23–60.

[5] HALLER, B. (1945). Verteilungsfunktionen und ihre Auszeichnung durch Funktional gleichungen. *Mitt. Verein. Schweiz. Versich.- math.* **45** No. 1, 97–163.

[6] PEARSON, KARL (1934). *Tables of the Incomplete Gamma Function.* Cambridge Univ. Press.

[7] ROBBINS, HERBERT (1948). The distribution of a definite quadratic form. *Ann. Math. Statist.* **19** 266–270.

IMPROVED BOUNDS ON A MEASURE OF SKEWNESS

BY KULENDRA N. MAJINDAR

Delhi University, India

In 1932, Hotelling and Solomons [2] proved that the absolute value of a certain measure of skewness for a population can not exceed 1. This result has been used by Madow [3] in his study of systematic sampling. The proof given by Hotelling and Solomons covers the case of a discrete random variable. In this note we extend and strengthen the inequality for any random variable with a positive standard deviation. Let X be a random variable with a positive standard deviation, M its median and $F(x)$ its cumulative distribution function. If the median is not uniquely defined, we will define it by $M = \frac{1}{2}\sup\{x: F(x) < \frac{1}{2}\} + \frac{1}{2}\inf\{x: F(x) > \frac{1}{2}\}$. The measure of skewness, S, considered here is the ratio of the difference between the mean and median to the standard deviation of X. With this definition we establish the following theorem.

THEOREM. *The measure of skewness S of a random variable X with a finite positive standard deviation satisfies the inequality*

$$|S| < 2(pq)^{\frac{1}{2}}/(p + q)^{\frac{1}{2}},$$

where $p = \Pr\{X > E(X)\}$ *and* $q = \Pr\{X < E(X)\}$.

Received June 26, 1961; revised February 26, 1962.