A LAGUERRE SERIES APPROXIMATION TO THE SAMPLING DISTRIBUTION OF THE VARIANCE

By J. ROY and M. L. TIKU
Indian Statistical Institute

SUMMARY. The first four terms of the Laguerre series expansion of the distribution of the variance of a sample from any population are worked out.

1. Introduction

An approximation to the sampling distribution of the variance in samples from any non-normal population was given by Gayen (1949). He started with a Gram-Charlier series expansion of the probability density function of the population and ignored all cumulants of the population above the fourth and also squares and higher powers of the fourth cumulant. An alternative approach is presented in this paper. The probability density function of the sample variance is expanded in terms of a Gamma density function and Laguerre polynomials and the coefficients of the first four terms are worked out explicitly in terms of population cumulants, of upto the eighth order. Gayen's expression agrees with this expansion to the order of approximation used by Gayen.

Laguerre polynomials. For $m > 0$, a Laguerre polynomial of degree r in x is defined as (see Szegö (1939), Chapter V)

$$L_r^{(m)}(x) = \sum_{t=0}^{r} C_{r,t}^{(m)} \frac{(-x)^t}{t!}, \qquad \ldots (1.1)$$

where $$C_{r,t}^{(m)} = \begin{cases} (m+t)(m+t+1)\ldots(m+r-1)/(r-t)!, & \text{for } t = 0, 1, \ldots, r-1 \\ 1 & \text{for } t = r \end{cases}$$

for $r = 0, 1, 2, \ldots$. In particular, the first four polynomials are :

$$L_0^{(m)}(x) = 1$$

$$L_1^{(m)}(x) = m-x$$

$$L_2^{(m)}(x) = \frac{1}{2!}\, m(m+1)-(m+1)x+\frac{x^2}{2!} \qquad \ldots (1.2)$$

$$L_3^{(m)}(x) = \frac{1}{3!} m(m+1)(m+2)-\frac{1}{2!}(m+1)(m+2)x+(m+2)\frac{x^2}{2!}-\frac{x^3}{3!}$$

$$L_4^{(m)}(x) = \frac{1}{4!}\, m(m+1)(m+2)(m+3)-\frac{1}{3!}(m+1)(m+2)(m+3)x$$
$$+\frac{1}{2!}(m+2)(m+3)\frac{x^2}{2!}-(m+3)\frac{x^3}{3!}+\frac{x^4}{4!}.$$

If we write

$$p_m(x) = \frac{1}{\Gamma(m)} e^{-x} x^{m-1} \quad \ldots \quad (1.3)$$

for the Gamma density function with mean m, the orthogonality property of Laguerre polynomials can be stated as :

$$\int_0^\infty L_r^{(m)}(x) L_s^{(m)}(x) p_m(x) dx = \begin{cases} 0 & \text{if } r \neq s \\ C_{r,0}^{(m)} & \text{if } r = s \end{cases} \quad \ldots \quad (1.4)$$

2. Approximate distribution of the corrected sum of squares

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population in which cumulants of all orders exist, and let the r-th cumulant be denoted by K_r, $r = 2, 3, \ldots$. The variance of the population will be alternatively denoted by $\sigma^2 = K_2$. Further, we shall write

$$\lambda_r = K_r K_2^{-\frac{1}{2}r} \; ; \quad r = 3, 4, \ \ldots \quad \ldots \quad (2.1)$$

Consider the sum of squares about the sample mean : $S^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$ where $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$. This is said to have $\nu = (n-1)$ degrees of freedom. We shall write

$$X = S^2/2\sigma^2 \quad \ldots \quad (2.2)$$

and try to derive an approximation for the probability density function of X.

We first note that the cumulants of the distribution of $s^2 = S^2/\nu$ has been worked out by Fisher (1928), and that the first few are tabulated in Kendall (1947). Using these, we get the first four central moments of X and these are listed below :

$$E(X) = m$$

$$\mu_2(X) = m + \frac{m^2}{2m+1} \lambda_4$$

$$\mu_3(X) = 2m + \frac{6m^2}{2m+1} \lambda_4 + \frac{m(2m-1)}{2m+1} \lambda_3^2 + \frac{m^3}{(2m+1)^2} \lambda_6 \quad \ldots \quad (2.3)$$

$$\mu_4(X) = 3m(m+2) + \frac{12m^2(m+3)}{2m+1} \lambda_4 + \frac{12m(2m-1)}{2m+1} \lambda_3^2 + \frac{12m^3}{(2m+1)^2} \lambda_6$$

$$+ \frac{m(3m^3+16m^2-2m+1)}{(2m+1)^2} \lambda_4^2 + \frac{8m^2(2m-1)}{(2m+1)^2} \lambda_5 \lambda_3 + \frac{m^4}{(2m+1)^3} \lambda_8$$

where we write for simplicity $\quad m = \frac{1}{2}(n-1). \quad \ldots \quad (2.4)$

Let us denote by $\phi_m(x)$ the probability density function of X. The quotient $\phi_m(x)/p_m(x)$ can be formally expanded in an infinite series in Laguerre polynomials as

$$\frac{\phi_m(x)}{p_m(x)} = \sum_{r=0}^{\infty} a_r^{(m)} L_r^{(m)}(x). \quad \ldots \quad (2.5)$$

Multiplying both sides of (2.5) by $L_r^{(m)}(x)\ p_m(x)$ and integrating over x from 0 to ∞, we get

$$a_r^{(m)} = \int_0^\infty L_r^{(m)}(x)\phi_m(x)dx/C_{r,0}^{(m)} \qquad \ldots \quad (2.6)$$

formally, by virtue of the orthogonal property (1.4) of Laguerre polynomials. For conditions of convergence of the formal expansion (2.5) see Szegö (1939), Chapter IX.

What we seek here is an approximation to $\phi_m(x)$ using only the first four terms in (2.5). Thus

$$\phi_m(x) \sim p_m(x)\ [\sum_{r=0}^{4} a_r^{(m)}\ L_r^{(m)}(x)] \qquad \ldots \quad (2.7)$$

where

$$a_r^{(m)} = \frac{E[L_r^{(m)}(X)]}{C_{r,0}^{(m)}},\ r = 0, 1, 2, 3, 4. \qquad \ldots \quad (2.8)$$

To evaluate $E[L_r^{(m)}(X)]$ for $r = 0, 1, 2, 3$ and 4, we note that writing $y = x-m$, the first four Laguerre polynomials given by (1.2) can be expressed in terms of y as :

$$L_1^{(m)} = -y$$

$$L_2^{(m)} = \frac{1}{2}y^2 - y - \frac{1}{2}m \qquad \ldots \quad (2.9)$$

$$L_3^{(m)} = -\frac{1}{6}y^3 + y^2 + \left(\frac{1}{2}m - 1\right)y - \frac{2}{3}m$$

$$L_4^{(m)} = \frac{1}{24}y^4 - \frac{1}{2}y^3 + \frac{1}{4}(6-m)\,y^2 + \left(\frac{7}{6}m - 1\right)y + \frac{1}{8}m(m-6).$$

Hence

$$E[L_1^{(m)}(X)] = 0$$

$$E[L_2^{(m)}(X)] = \frac{1}{2}\mu_2 - \frac{1}{2}m \qquad \ldots \quad (2.10)$$

$$E[L_3^{(m)}(X)] = -\frac{1}{6}\mu_3 + \mu_2 - \frac{2}{3}m$$

$$E[L_4^{(m)}(X)] = \frac{1}{24}\mu_4 - \frac{1}{2}\mu_3 + \frac{1}{4}(6-m)\mu_2 + \frac{1}{8}m(m-6)$$

where μ_2, μ_3, μ_4 are respectively the second, third and fourth central moments of X given by (2.3). Using (2.8) and (2.10), we finally get

$$a_0^{(m)} = 1$$

$$a_1^{(m)} = 0$$

$$a_2^{(m)} = \frac{m}{(2m+1)(m+1)}\ \lambda_4 \qquad \ldots \quad (2.11)$$

$$a_3^{(m)} = -\frac{1}{(2m+1)(m+1)(m+2)}\left[\frac{m^3}{2m+1}\ \lambda_6 + (2m-1)\ \lambda_3^2\right]$$

$$a_4^{(m)} = \frac{1}{(2m+1)(m+1)(m+2)(m+3)}\left[\frac{m^3}{(2m+1)^2}\lambda_8 + \frac{8m(2m-1)}{2m+1}\ \lambda_5\lambda_3 + \frac{3m^3+16m^2-2m+1}{(2m+1)}\ \lambda_4^2\right]$$

11

The probability density function of the sample variance $Z = \frac{S^2}{n-1} = \frac{\sigma^2 X}{m}$ is then obtained as

$$m\sigma^{-2} p_m(m\sigma^{-2}z) \sum_{r=0}^{4} a_r^{(m)} L_r^{(m)}(m\sigma^{-2}z) \qquad \dots \quad (2.12)$$

If terms involving λ_r for $r > 4$ and λ_3^2 are ignored, this agrees with Gayen's (1949) formula (3.1).

REFERENCES

GAYEN, A. K. (1949): The distribution of Student's *t* in random samples of any size from non-normal universes. *Biometrika*, **36**, 353-369.

SZEGÖ, G. (1939): Orthogonal polynomials (*Amer. Math. Soc. Colloquium Publication* Vol. XXIII).

KENDALL, M. G. (1947): *Advanced Theory of Statistics*, Vol. I, Third edition, Charles Griffin & Company Ltd.

FISHER, R. A. (1928): Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc.*, **30**, 199–238.

Paper received : December, 1961.