ON THE ASYMPTOTIC DISTRIBUTIONS OF CERTAIN STATISTICS USED IN TESTING THE INDEPENDENCE BETWEEN SUCCESSIVE OBSERVATIONS FROM A NORMAL POPULATION

By P. L. Hsu

Columbia University

1. The statistics to be considered here have the general expression

$$T = \frac{Q}{S}, \qquad Q = \sum_{i=1}^{N} a_{ij}(x_i - \bar{x})(x_j - \bar{x}), \qquad S = \sum_{i=1}^{N} (x_i - \bar{x})^2,$$

where $(x_1, \cdots, x_N)$ is a sample from a normal population whose mean and variance can evidently be assumed to be 0 and 1 respectively.[1] The purpose of this note is to study the asymptotic distribution of T assuming that the x_i are independent. The whole work may be regarded as a straightforward application of Cramér's theory of asymptotic expansion (see [1], pp. 69–88).

If $A = [a_{ij}]$ and γ is the row vector $N^{-\frac{1}{2}}[1, 1, \cdots, 1, 1]$ the quadratic form Q has the matrix $(I - \gamma'\gamma)A(I - \gamma'\gamma)$. The latent roots of this matrix, which are also the latent roots of $A(I - \gamma'\gamma)^2 = A(I - \gamma'\gamma)$, will be denoted by $0, \lambda_1, \cdots, \lambda_n$, with $n = N - 1$. Then Q and S can be simultaneously diagonalized (by a rotation of the N-dimensional space), so that

$$Q = \sum_{r=1}^{N} \lambda_r y_r^2, \qquad S = \sum_{r=1}^{n} y_r^2,$$

where the y_r are again independently and normally distributed with zero mean and unit variance.

We shall make the following assumptions

(a) $|\lambda_r| \leq 1$ for all r.

(b) There is a positive number c independent of n such that

$$\sum_{r=1}^{n} (\lambda_r - \bar{\lambda})^2 > cn, \quad \text{where} \quad \bar{\lambda} = \frac{1}{n}\sum_{r=1}^{n} \lambda_r.$$

Write

$$z = \frac{\sqrt{2\sum_{r=1}^{n}(\lambda_r - \bar{\lambda})^2 x}}{\sqrt{n^2 - 2nx^2}}, \qquad s_m(x) = \sum_{r=1}^{n} (\lambda_r - \bar{\lambda} - z)^m,$$

$$X_r = (\lambda_r - \bar{\lambda} - z)(y_r^2 - 1), \qquad G(x) = Pr\{T \leq \bar{\lambda} + z\}.$$

[1] The exact and the approximate distribution of such statistics were a recent subject of study by a number of statisticians. See W. J. Dixon, "Further contributions to the problem of serial correlation," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 119–144. Further references are listed in Dixon's paper.

Then it can easily be verified that

$$G(x) = Pr\left\{\frac{\sum_{i=1}^{r} X_r}{\sqrt{2s_2(x)}} \leq x\right\}.$$

This expression of $G(x)$ shows that the application of Cramér's expansion is at hand, since $E(X_r) = 0$ and $2s_2(x)$ is the variance of ΣX_r. Let ρ_{kn} and T_{kn} stand for the same quantities as defined in Cramér's work (see [1], pp. 70–71). Since moments of all order of X_r exist, we may use $2k + 2$ in place of k. We have

$$\rho_{2k+2,n} = \frac{\frac{1}{n} m_k s_{2k+2}(x)}{\left(\frac{2}{n} s_2(x)\right)^{k+1}}, \qquad T_{2k+2,n} = \frac{\sqrt{n}}{4\rho_{2k+2,n}^{3/2k+2}},$$

where $m_k = E(y^2 - 1)^{2k+2}$ and y is a normal variate with mean 0 and variance 1.

By virtue of assumption (a) $| T | \leq 1$. Therefore we may confine ourselves to the range of values for which $| \bar{\lambda} + z | \leq 1$. Then $| \lambda_r - \bar{\lambda} - z | \leq 2$. Also, by assumption (b), $s_2(x) \geq \Sigma(\lambda_r - \bar{\lambda})^2 > cn$. Hence $\rho_{2k+2,n}$, and in consequence $\sqrt{n}T_{2k+2,n}^{-1}$, are less than some constant independent of n and x. The remainder of Cramer s expansion, if it is justifiable, will therefore be less than Mn^{-k}, where M is independent of n and x. The justification consists in verifying that the following condition is satisfied: if $f_r(t)$ is the characteristic function of X_r and A is any positive number, then

$$\text{l.u.b.} \prod_{r=1}^{n} |f_r(t)| \quad \text{for} \quad |t| > \frac{T_{2k+2,n}}{\sqrt{2s_2(x)}}$$

is less than $M_1 T_{2k+2,n}^{-A}$, where M_1 is independent of n and x (see [1], p. 85). Since $T_{2k+2,n} \leq \frac{1}{4}\sqrt{n}$ [2] and $s_2(x) > c\sqrt{n}$, it is sufficient to show that, if a and A are any positive numbers and if

$$U = \text{l.u.b.} \prod_{r=1}^{n} |f_r(t)| \quad \text{for} \quad |t| > a,$$

then $U \leq M_2 n^{-A}$, where M_2 is independent of n and x. Now

$$|f_r(t)| = \{1 + 4t^2(\lambda_r - \bar{\lambda} - z)^2\}^{-\frac{1}{4}}$$

whence

$$U = \prod_{r=1}^{n} \{1 + 4a^2(\lambda_r - \bar{\lambda} - z)^2\}^{-\frac{1}{4}}.$$

Let μ be the number of λ_r for which $(\lambda_r - \bar{\lambda} - z)^2 < \frac{1}{2}c$. Then $cn < s_2(x) \leq \frac{1}{2}c(n - \mu) + 4\mu$; hence $cn < (8 - c)\mu$ and

$$U \leq (1 + 2a^2c)^{-\frac{1}{4}\mu} < (1 + 2a^2c)^{-(cn/4(8-c))}$$

This shows that the desired condition on U is satisfied, and that therefore Cramér's procedure can be adopted.

[2] This follows from the fact that $P_{2k+2,n} > 1$. Cf. Cramér, [1], p. 70.

Wherever Cramér's asymptotic expansion is valid, the terms in the expansion are most conveniently obtained with the help of Cornish and Fisher's symbolic expression (see [2]):

$$e^{-(1/3!)\gamma_3(d^3/dx^3)+(1/4!)\gamma_4(d^4/dx^4)-\cdots}\Phi(x),$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}y^2}\,dy$$

and γ_j is the jth semi-invariant of the random variable whose distribution is under asymptotic expansion. In the present case we have

$$\frac{\gamma_j}{j!} = \frac{\beta_j(x)}{n^{\frac{1}{2}(j-2)}},$$

where

$$\beta_j(x) = \frac{2^{\frac{1}{2}(j-2)}}{j}\,\frac{\frac{1}{n}s_j(x)}{\left(\frac{1}{n}s_2(x)\right)^{\frac{1}{2}j}}.$$

Hence we may express our result as follows:

$$G(x) = \exp\left[\sum_{j=3}^{2k+1}\frac{(-1)^j\beta_j(x)}{n^{\frac{1}{2}(j-2)}}\left(\frac{d}{dx}\right)^j\right]\Phi(x) + R_k(x), \tag{1}$$

where $| R_k(x) | \leq Mn^{-k}$, and M is independent of n and x. The symbolic exponential in (1) is to be expanded as far as and including the term in $n^{-\frac{1}{2}(2k-1)}$.

2. Let us apply the result (1) to the following three statistics: $T_\alpha = Q_\alpha/S$, $(\alpha = 1, 2, 3)$, where

$$Q_1 = \sum_{i=1}^{N}(x_i - \bar{x})(x_{i+1} - \bar{x}) \quad \text{with} \quad x_{N+1} = x_1,$$

$$Q_2 = \tfrac{1}{2}(x_1 - \bar{x})^2 + \tfrac{1}{2}(x_N - \bar{x})^2 + \sum_{i=1}^{N-1}(x_i - \bar{x})(x_{i+1} - \bar{x}),$$

$$Q_3 = \sum_{i=1}^{N-1}(x_i - \bar{x})(x_{i+1} - \bar{x}).$$

T_2 is simply related with $T^* = Q^*/S$, where

$$Q^* = \sum_{i=1}^{N-1}(x_i - x_{i+1})^2;$$

for we have $Q_2 = S - \frac{1}{2}Q^*$, whence $T_2 = 1 - \frac{1}{2}T^*$. We shall write $\lambda_r^{(\alpha)}$ for the λ's corresponding to Q_α, and

$$b_{m\alpha} = \sum_{r=1}^{n}(\lambda_r^{(\alpha)})^m, \qquad (\alpha = 1, 2, 3).$$

(i) For Q_1 we have $\lambda_r^{(1)} = \cos \frac{2\pi r}{N}$ (see [3]). Since

$$\cos^m \theta = \frac{1}{2^m} (e^{i\theta} + e^{-i\theta})^m = \frac{1}{2^m} \sum_{j=0}^{m} \binom{m}{j} e^{i(2j-m)\theta},$$

we have

$$b_{m1} = \frac{1}{2^m} \sum_{j=0}^{m} \binom{m}{j} \sum_{r=1}^{n} \xi^r, \quad \text{where} \quad \xi = e^{2\pi(2j-m)i/N}.$$

If $m < n$, then

$$\sum_{r=1}^{n} \xi^r = -1 \quad \text{if} \quad j \neq \tfrac{1}{2}m, \qquad = n \quad \text{if} \quad j = \tfrac{1}{2}m.$$

Hence, for $m < n$, $b_{m1} = -1$ if m is odd, $b_{m1} = \frac{N}{2^m} \binom{m}{\frac{1}{2}m} - 1$ if m is even.

In particular

$$\bar{\lambda}^{(1)} = -\frac{1}{n}, \qquad \sum_{r=1}^{n} (\lambda_r^{(1)} - \bar{\lambda}^{(1)})^2 = \frac{n^2 - n - 2}{2n} > 0.4n \quad \text{if} \quad n \geq 7.$$

Hence assumptions (a) and (b) are true (for $n \geq 7$). The $s_j(x)$ are conveniently computed with the help of b_{m1}. The $\beta_j(x)$ are then computed to yield the terms in (1).

(ii) The λ's corresponding to Q^* are $4 \sin^2 \frac{r\pi}{2N}$ (see [4]). Hence

$$\lambda_r^{(2)} = \cos \frac{r\pi}{N}.$$

By a computation similar to that in (i) we easily obtain $b_{m2} = \frac{N}{2^m} \binom{m}{\frac{1}{2}m} - 1$ for even m and $b_{m2} = 0$ for odd m, provided $m < 2n$. In particular, $\bar{\lambda}^{(2)} = 0$, $\Sigma(\lambda_r^{(2)} - \bar{\lambda}^{(2)})^2 = \frac{n-1}{2} \geq \cdot 4n$ for $n \geq 5$. Hence assumptions (a) and (b) are true (for $n \geq 5$).

(iii) In the case of Q_3 the matrix A is

$$A = \begin{Vmatrix} 0 & \frac{1}{2} & & & & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & & & \\ & \frac{1}{2} & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & 0 & \frac{1}{2} \\ 0 & & & & \frac{1}{2} & 0 \end{Vmatrix}$$

whose latent roots are $\cos \pi t/(N + 1)$, $(t = 1, \cdots, N)$ (see [5]), all less than or equal to unity in absolute value. It follows that the same is true for the $\lambda_r^{(3)}$.

Hence assumption (a) is true. Unlike the two previous cases, there is no simple expression for b_{m3}. With the help of the formula

$$b_{m3} = \operatorname{tr} \{A(I - \gamma'\gamma)\}^m$$

we may compute b_{m3} for small values of m. Thus

$$b_{13} = -\frac{n}{n+1}$$

$$b_{23} = \frac{n}{2} - \frac{2n-1}{n+1} + \frac{n^2}{(n+1)^2}$$

$$b_{33} = -\frac{3(n-1)}{n+1} + \frac{3n(2n-1)}{2(n+1)^2} - \frac{n^3}{(n+1)^3}$$

$$b_{43} = \frac{3n-2}{8} - \frac{8n-11}{2(n+1)} + \frac{4n(n-1)}{(n+1)^2} + \frac{(2n-1)^2}{2(n+1)^2} - \frac{2n^2(2n-1)}{(n+1)^3} + \frac{n^4}{(n+1)^4}$$

$$b_{53} = -\frac{5(4n-7)}{4(n+1)} + \frac{5n(8n-11)}{8(n+1)^2} + \frac{5(2n-1)(n-1)}{2(n+1)^2} - \frac{5n^2(n-1)}{(n+1)^3} - \frac{5n(2n-1)^2}{4(n+1)^3} + \frac{5n^3(2n-1)}{2(n+1)^5} - \frac{n^5}{(n+1)^5}$$

$$\overline{\lambda^{(3)}} = -\frac{1}{n+1}, \sum_{r=1}^{n} (\lambda_r^{(3)} - \overline{\lambda^{(3)}})^2 = \frac{n}{2} - \frac{2n-1}{n+1} + \frac{n^2-n}{(n+1)^2} > 0.4n \text{ for } n \geq 10.$$

Hence assumption (b) is true (for $n \geq 10$). Using these values of b_{m3} we may compute $\beta_3(x)$, $\beta_4(x)$ and $\beta_5(x)$. By (1) we have

$$G(x) = \Phi(x) - \frac{1}{n^{\frac{1}{2}}} \beta_3(x)\Phi^{(3)}(x) + \frac{1}{n} (\beta_4(x)\Phi^{(4)}(x) + \tfrac{1}{2}\beta_3^2(x)\Phi^{(6)}(x)) - \frac{1}{n^{\frac{3}{2}}} (\beta_5(x)\Phi^{(5)}(x) - \beta_3(x)\beta_4(x)\Phi^{(7)}(x) + \tfrac{1}{6}\beta_3^3(x)\Phi^{(9)}(x)) + R(x),$$

where $|R(x)| \leq Mn^{-2}$ and M is independent of n and x.

REFERENCES

[1] H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tract No. 36, 1937.

[2] E. A. Cornish and R. A. Fisher, *Revue de l'Institut International de Statistique*, 1937.

[3] R. L. Anderson, "Distribution of the serial correlation coefficient", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 1–13.

[4] T. Koopmans, "Serial correlation and quadratic forms in normal variables", *Annals of Math. Stat.* Vol. 13 (1942), pp. 14–33.

[5] J. von Neumann, "A further remark concerning the distribution of the ratio of the mean square successive difference to the variance", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 86–88.