DISTRIBUTION OF THE SERIAL CORRELATION COEFFICIENT

By R. L. Anderson

North Carolina State College

1. Introduction. The problem of serial correlation was brought to the attention of statisticians by Yule in 1921 [9]. Both Yule and Bartlett [2] have shown that the ordinary tests of significance are invalidated if successive observations are not independent of one another. The serial correlation coefficient has been introduced as a measure of the relationship between successive values of a variable ordered in time or space. Interest in the serial correlation problem was stimulated further by the new concepts of time series analysis discussed by Wold [8].

We shall define the serial correlation coefficient for lag L and N observations to be

$$_LR_N = \frac{_LC_N}{V_N} = \frac{X_1X_{L+1} + X_2X_{L+2} + \cdots + X_NX_L - (\Sigma X_i)^2/N}{\Sigma X_i^2 - (\Sigma X_i)^2/N},$$

where C and V are the covariance and variance respectively and the X's are considered to be independently normally distributed about the same mean with unit variance.[1] If the population variance were known a priori, the variates could be transformed so that they would have unit variance; under such an unusual circumstance, the only distribution required would be that of the serial covariance. Tintner has given a test of significance for the serial covariance [6] and for the correlation coefficient [7] by using a method of selected items. The author has presented the distribution of the serial covariance and of the serial correlation coefficient not corrected for the mean in a recent doctoral thesis [1]. The distributions of $_LR_N$ not corrected for the mean will be mentioned in the sections which follow.

2. Small sample distributions for lag 1. W. G. Cochran has suggested that we use a result given in his article on quadratic forms to derive the distributions of the serial correlation coefficient for small samples [3]. If $X_1, X_2, \cdots, X_N$ are independently normally distributed with variance 1 and mean 0, then

> "Every quadratic form $\Sigma a_{ij}X_iX_j$ is distributed like $\sum_{k=1}^{r} \lambda_k u_k$, where r is the rank of the matrix, A, of the quadratic form, the u's are independently distributed as χ^2, each with 1 d.f., and the λ's are the non-zero latent roots of the characteristic equation of A" [3, p. 179].

If each λ_i appears k_i times as a latent root, u_i will be distributed as χ^2 with k_i degrees of freedom.

[1] This circular definition of the serial correlation coefficient was suggested by H. Hotelling.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org