

DISTRIBUTION OF THE SERIAL CORRELATION COEFFICIENT
IN A CIRCULARLY CORRELATED UNIVERSE¹

BY R. B. LEIPNIK

Cowles Commission for Research in Economics

1. **Summary.** It is desired to find an approximate distribution of simple form for the statistic $\bar{r} = \frac{x_1 x_2 + \dots + x_T x_1}{x_1^2 + \dots + x_T^2}$ (\bar{r} is an estimate of the serial correlation coefficient ρ in a circular universe) in the case that $\rho \neq 0$ in the universe. Such a distribution is obtained by smoothing the joint characteristic function of the numerator and denominator of the expression for \bar{r} . The first two moments are calculated; from these \bar{r} is seen to be a consistent estimate of ρ . A graph of this distribution for sample size $T = 20$ and various values of ρ is given.

In addition, an approximate distribution for $p = x_1^2 + \dots + x_T^2$ is derived which reduces to the exact (χ^2 -) distribution if $\rho = 0$. From a formula which yields all moments, it is concluded that, at least up to the degree of approximation attained, p/T is an unbiased and consistent estimate of σ^2 .

2. Several writers have investigated the temporally homogeneous stochastic process defined by

$$(1) \quad x_t - \rho x_{t-1} = z_t, \quad t = 1, 2, \dots, T, \quad |\rho| < 1,$$

where the z_t are unobservable disturbances, normally and independently distributed with mean zero and variance σ^2 , the x_t are observed variates, and the "first observation" x_0 has a normal distribution with mean zero and such a variance σ_z^2 that all later observations have the same variance. Thus we have

$$(2) \quad \sigma_z^2 = \frac{\sigma^2}{1 - \rho^2}$$

and the joint distribution of a sample of $T + 1$ successive values is

$$(3) \quad g(x_0, x_1, \dots, x_T) = \frac{(1 - \rho^2)^{\frac{1}{2}}}{(2\pi\sigma^2)^{T/2+1/2}} \cdot \exp \left[-\frac{1}{2\sigma^2} \{x_0^2 + x_T^2 - 2\rho(x_0 x_1 + \dots + x_{T-1} x_T) + (1 + \rho^2)(x_1^2 + \dots + x_{T-1}^2)\} \right].$$

Koopmans ([1], formula 96), by smoothing characteristic values, has obtained an approximation to the distribution of the serial correlation coefficient r for the case $\rho = 0$, where

$$(4) \quad r = \frac{x_0 x_1 + \dots + x_{T-1} x_T}{x_0^2 + \dots + x_T^2}.$$

¹ Cowles Commission Papers, New Series, No. 21.

This result is expressed in the form of a definite integral whose evaluation has not so far been effected.

By considering the related circular stochastic process, where x_0 is defined to be the same observation as x_T , great simplification is obtained. Here the joint distribution of x_1, x_2, \dots, x_T is

$$f(x_1, x_2, \dots, x_T) = \frac{\lambda(\rho)}{(2\pi\sigma^2)^{T/2}} \exp \left[-\frac{1}{2\sigma^2(1-\rho^2)} \right. \\ \left. \{ (1+\rho^2)(x_1^2 + \dots + x_T^2) - 2\rho(x_1x_2 + \dots + x_Tx_1) \} \right] \\ \lambda(\rho) = \frac{1-\rho^T}{(1-\rho^2)^{T/2}}.$$

By smoothing characteristic values, Koopmans ([1], formula 92) found a definite integral and Dixon ([2], 3.22) an explicit expression for an approximate distribution of the circular serial correlation coefficient \bar{r} , for the case $\rho = 0$, where

$$\bar{r} = \frac{x_1x_2 + \dots + x_Tx_1}{x_1^2 + \dots + x_T^2}.$$

Dixon's distribution $\bar{K}_0(\bar{r})$ has the simple form

$$\bar{K}_0(\bar{r}) = \frac{\Gamma\left(\frac{T}{2} + 1\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{T}{2} + 2\right)} (1 - \bar{r}^2)^{T/2-1}.$$

Rubin [3] proved these results to be equivalent. On the other hand, R. L. Anderson [4] obtained the exact distribution of \bar{r} in the case $\rho = 0$. Madow [5] extended this result to the case $\rho \neq 0$, using a property of sufficient statistics also noted by Koopmans ([1], p. 17) in connection with the non-circular problem.

It would, however, be difficult to find percentile points or moments from Madow's exact distribution. An approximate distribution of \bar{r} for $\rho \neq 0$, together with its moments, analogous to Dixon-Koopmans' for $\rho = 0$, should therefore be of interest. The purpose of this paper is to obtain such a distribution from the circular universe (5). The statistic \bar{r} is shown to be a consistent estimate of ρ within the limits imposed by the approximation. In addition, an approximate distribution for $p = x_1^2 + \dots + x_T^2$ in the case $\rho \neq 0$ (which reduces to the exact chi-squared distribution when $\rho = 0$) is derived, together with all of its moments.

3. We begin by asking about an approximate joint distribution of p and \bar{q} defined by

$$p = x_1^2 + \dots + x_T^2 \\ \bar{q} = x_1x_2 + \dots + x_Tx_1.$$

Defining $\phi(u, v)$ as the expectation of $\exp[i(up + v\bar{q})]$, we have

$$(9) \quad \phi(u, v) = \frac{\lambda(\rho)}{(2\pi\sigma^2)^{T/2}} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp \left[-\frac{1}{2\sigma^2} \left\{ \left(\frac{1 + \rho^2}{1 - \rho^2} - 2i\sigma^2 u \right) p - 2 \left(\frac{\rho}{1 - \rho^2} + i\sigma^2 v \right) \bar{q} \right\} \right] dx_1 \cdots dx_T.$$

On integration, we find

$$(10) \quad \phi(u, v) = \lambda(\rho) [A(u, v)]^{-1/2}$$

where $A(u, v)$ is the determinant of the matrix associated with the quadratic form within the curly brackets in (9). $A(u, v)$ is a circulant; its value as determined from the circulant formula ([2], p. 123) is

$$(11) \quad A(u, v) = \prod_{t=1}^T \left(y - 2z \cos \frac{2\pi t}{T} \right)$$

where y and z are defined by

$$(12) \quad y = \frac{1 + \rho^2}{1 - \rho^2} - 2i\sigma^2 u$$

$$z = \frac{\rho}{1 - \rho^2} + i\sigma^2 v.$$

To get an approximation $\tilde{A}(u, v)$ to $A(u, v)$ we smooth $\log A(u, v)$ by Koopmans' method. We have

$$(13) \quad \log A(u, v) = \sum_{t=1}^T \log \left(y - 2z \cos \frac{2\pi t}{T} \right).$$

We define $\tilde{A}(u, v)$ through

$$(14) \quad \log \tilde{A}(u, v) = \int_0^T \log \left(y - 2z \cos \frac{2\pi t}{T} \right) dt$$

in which the summation in (13) is replaced by integration. The integral in (14) is easily evaluated ([6], p. 65) giving

$$(15) \quad \tilde{A}(u, v) = \left(\frac{y + \sqrt{y^2 - 4z^2}}{2} \right)^T.$$

Incidentally, had we used $\bar{q}_L = x_1 x_{L+1} + \cdots + x_T x_{T+L}$ in place of $\bar{q}_1 = \bar{q}$ in (9), we would have obtained the same expression (15) for $\tilde{A}(u, v)$.

Setting $\tilde{\phi}(u, v) = \tilde{\lambda}(\rho) [\tilde{A}(u, v)]^{-1/2}$ we may determine $\tilde{\lambda}(\rho)$ by the requirement $\tilde{\phi}(0, 0) = 1$. A simple calculation yields the result $\tilde{\lambda}(\rho) = (1 - \rho^2)^{-(T/2)}$. (Note that $\frac{\lambda(\rho)}{\tilde{\lambda}(\rho)} = 1 - \rho^T$ is close to 1 for large values of T). Our result for $\tilde{\phi}(u, v)$ appears as

$$(16) \quad \tilde{\phi}(u, v) = \tilde{\lambda}(\rho) \left(\frac{y + \sqrt{y^2 - 4z^2}}{2} \right)^{-(T/2)}.$$

The approximate joint distribution of p and \bar{q} may be written as the double Fourier integral

$$(17) \quad \bar{D}(p, \bar{q}) = \frac{\tilde{\lambda}(\rho)}{4\pi^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp[-i(Up + v\bar{q})] \left(y + \frac{\sqrt{y^2 - 4z^2}}{2} \right)^{-T/2} du dv$$

which we evaluate ([7], 576.3, 914.3) by changing integration variables from u, v to y, z and integrating out y and z successively. We obtain finally

$$(18) \quad \bar{D}(p, \bar{q}) = \frac{T}{2} \cdot \frac{[2\sigma^2(1 - \rho^2)]^{-T/2}}{\Gamma(\frac{1}{2})\Gamma(\frac{T}{2} + \frac{1}{2})} p^{-(T/2)-1} (p^2 - \bar{q}^2)^{T/2-1} \cdot \exp\left[-\frac{1}{2\sigma^2(1 - \rho^2)} \{(1 + \rho^2)p - 2\rho\bar{q}\}\right].$$

Changing variables from $p, \bar{q} = p\bar{r}$ to p, \bar{r} , we obtain for $\bar{F}(p, \bar{r})$, the approximate joint distribution of p and \bar{r} , the expression

$$(19) \quad \bar{F}(p, \bar{r}) = \frac{T}{2} \cdot \frac{[2\sigma^2(1 - \rho^2)]^{-T/2}}{\Gamma(\frac{1}{2})\Gamma(\frac{T}{2} + \frac{1}{2})} p^{T/2-1} (1 - \bar{r}^2)^{T/2-1} \cdot \exp\left[-\frac{p}{2\sigma^2(1 - \rho^2)} \{1 + \rho^2 - 2\rho\bar{r}\}\right].$$

We could also have derived (19), following Madow, by noting that for $\rho = 0$, p and \bar{r} are independently distributed, p having the chi-squared distribution and \bar{r} having approximately the Dixon distribution (7), and that p and \bar{r} are sufficient statistics for the estimation of ρ and σ^2 .

4. The approximate marginal distribution $\tilde{R}_\rho(\bar{r})$ of \bar{r} is obtained by an easy integration from (19)

$$(20) \quad \tilde{R}_\rho(\bar{r}) = \int_0^\infty \bar{F}(p, \bar{r}) dp = \frac{T [2\sigma^2(1 - \rho^2)]^{-T/2}}{\Gamma(\frac{1}{2})\Gamma(\frac{T}{2} + \frac{1}{2})} (1 - \bar{r}^2)^{T/2-1} \cdot \int_0^\infty p^{T/2-1} \exp\left[-\frac{p}{2\sigma^2(1 - \rho^2)} \{1 + \rho^2 - 2\rho\bar{r}\}\right] dp, \\ \tilde{R}_\rho(\bar{r}) = \frac{\Gamma(\frac{T}{2} + 1)}{\Gamma(\frac{1}{2})\Gamma(\frac{T}{2} + \frac{1}{2})} (1 - \bar{r}^2)^{T/2-1} (1 + \rho^2 - 2\rho\bar{r})^{-T/2}.$$

Our notation is consistent since $\tilde{R}_\rho(\bar{r})$ indeed reduces to the Dixon distribution for $\rho = 0$. $\tilde{R}_\rho(\bar{r})$ has a maximum when

$$\bar{r} = \bar{r}_{\max} = \frac{1}{2\rho(T - 2)} \left\{ (1 + \rho^2)(T - 1) - \sqrt{T(T - 2)(1 - \rho^2)^2 + (1 + \rho^2)^2} \right\}.$$

A little manipulation shows that $1 > |\bar{r}_{\max}| > |\rho|$ and that $\bar{r}_{\max} = \rho$ asymptotically. A graph (Fig. 1) of $\tilde{R}_\rho(\bar{r})$ for $T = 20, \rho = 0, .2, .5, .7, .9$ is appended from which it is seen that for $|\rho|$ near 1, the distribution becomes highly concentrated about \bar{r}_{\max} . On differentiating $\tilde{R}_\rho(\bar{r})$ with respect to ρ and eliminating, the envelope of the $\tilde{R}_\rho(\bar{r})$ is seen to be

$$\frac{\Gamma\left(\frac{T}{2} + 1\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{T}{2} + \frac{1}{2}\right)} (1 - \bar{r}^2)^{-\frac{1}{2}}.$$

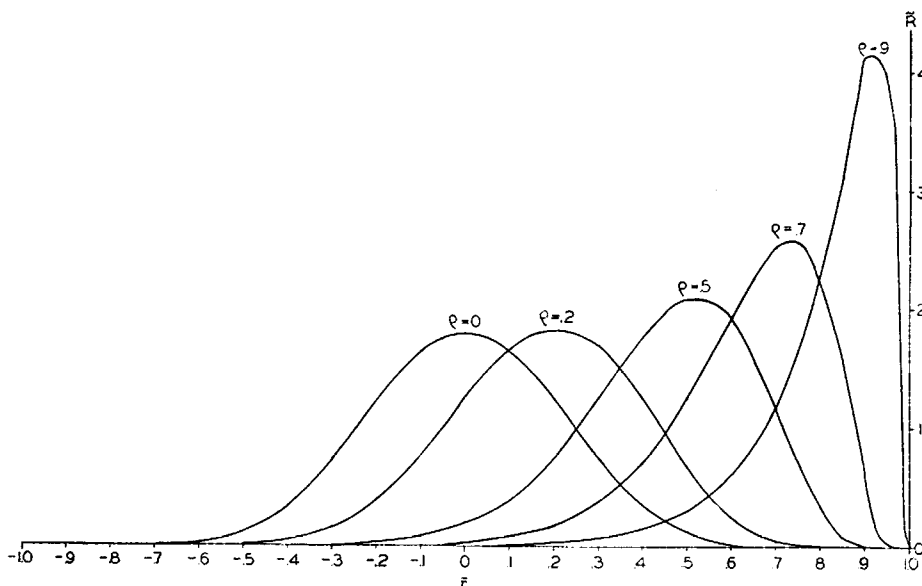


Fig. 1. Graph of the Distribution of the Serial Correlation Coefficient in a Circular Universe, for $T = 20$

5. Before evaluating the moments of $\tilde{R}_\rho(\bar{r})$ we will pause to obtain the approximate marginal distribution $\tilde{P}_\rho(p)$ of p , and its moments. We write

$$(21) \quad \tilde{P}_\rho(p) = \int_{-1}^{+1} \tilde{F}(p, \bar{r}) d\bar{r} = \frac{T}{2} \cdot \frac{[2\sigma^2(1 - \rho^2)]^{-T/2}}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{T}{2} + \frac{1}{2}\right)} \\ \cdot p^{T/2-1} \exp\left[-\frac{p}{2\sigma^2} \left(\frac{1 + \rho^2}{1 - \rho^2}\right)\right] \cdot \int_{-1}^{+1} (1 - \bar{r}^2)^{T/2-1} \exp\left[\frac{\rho p \bar{r}}{\sigma^2(1 - \rho^2)}\right] d\bar{r}.$$

If we define $I_\nu(z)$, the Bessel function of order ν and purely imaginary argument by

$$(22) \quad I_\nu(z) = \sum_{n=0}^{\infty} \frac{\left(\frac{z}{2}\right)^{\nu+2n}}{n! \Gamma(\nu+n+1)},$$

we obtain ([8], p. 79), if $\rho \neq 0$

$$(23) \quad \tilde{P}_\rho(p) = \frac{T}{2} \rho^{-T/2} p^{-1} \exp \left[-\frac{p}{2\sigma^2} \left(\frac{1+\rho^2}{1-\rho^2} \right) \right] I_{T/2} \left(\frac{\rho p}{\sigma^2(1-\rho^2)} \right),$$

and if $\rho = 0$

$$(24) \quad \tilde{P}_0(p) = \frac{(2\sigma^2)^{-T/2}}{\Gamma\left(\frac{T}{2}\right)} p^{T/2-1} \exp \left[-\frac{p}{2\sigma^2} \right],$$

on performing the integration indicated in (21). $\tilde{P}_0(p)$ coincides with the exact distribution $P_0(p)$. An expression covering all moments of $\tilde{P}_\rho(p)$ is obtained from (16) by setting $\nu = 0$, differentiating, and setting $u = 0$. We have

$$(25) \quad \tilde{\phi}(u, 0) = \tilde{\lambda}(\rho) \left(\frac{y + \sqrt{y^2 - \left[\frac{2\rho}{1-\rho^2} \right]^2}}{2} \right)^{-T/2},$$

hence

$$(26) \quad \tilde{E}[p^k] = i^{-k} \frac{d^k}{du^k} \tilde{\phi}(u, 0) \Big|_{u=0} = (-2\sigma^2)^k (1-\rho^2)^{-T/2} \cdot \frac{d^k}{dy^k} \left(\frac{y + \sqrt{y^2 - \left[\frac{2\rho}{1-\rho^2} \right]^2}}{2} \right)^{-T/2} \Big|_{y=(1+\rho^2)/(1-\rho^2)}.$$

From (26), we readily find

$$(27) \quad \tilde{E}[p] = T\sigma^2, \quad \tilde{E} \left[\frac{p}{T} \right] = \sigma^2$$

$$(28) \quad \tilde{E}[p^2] = (T\sigma^2)^2 + 2T\sigma^4 \left(\frac{1+\rho^2}{1-\rho^2} \right)$$

$$\tilde{\sigma}_p^2 = 2T\sigma^4 \left(\frac{1+\rho^2}{1-\rho^2} \right), \quad \tilde{\sigma}_{p/T}^2 = \frac{2\sigma^4}{T} \left(\frac{1+\rho^2}{1-\rho^2} \right).$$

Thus the unbiased character of p/T as an estimate of σ^2 is reflected in the approximate distribution, while (28), which shows that $\lim_{T \rightarrow \infty} \tilde{\sigma}_{p/T}^2 = 0$, indicates that consistency is also reflected.

6. We now calculate the moments of $\hat{R}_\rho(\bar{r})$. Interchanging the order of integration in the expression for $\tilde{E}[\bar{r}^k]$ is justified by the uniform convergence, so we have

$$\begin{aligned}
\tilde{E}[\bar{r}^k] &= \int_{-1}^{+1} \bar{r}^k \left[\int_0^\infty \bar{F}[p, \bar{r}] dp \right] d\bar{r} = \int_0^\infty \left[\int_{-1}^{+1} \bar{r}^k \bar{F}(p, \bar{r}) d\bar{r} \right] dp \\
(29) \quad &= \frac{T}{2} \frac{[2\sigma^2(1 - \rho^2)]^{-T/2}}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{T}{2} + \frac{1}{2}\right)} \int_0^\infty p^{T/2-1} \\
&\cdot \exp\left[-\frac{p}{2\sigma^2} \left(\frac{1 + \rho^2}{1 - \rho^2}\right)\right] \left\{ \int_{-1}^{+1} \bar{r}^k (1 - \bar{r}^2)^{T/2-1/2} \exp(m\bar{r}) d\bar{r} \right\} dp
\end{aligned}$$

where m is defined by

$$(30) \quad m = \frac{\rho p}{\sigma^2(1 - \rho^2)}.$$

Defining $G(m)$ by

$$(31) \quad G(m) = \int_{-1}^{+1} (1 - \bar{r}^2)^{T/2-1/2} \exp(m\bar{r}) d\bar{r}$$

we have ([8], p. 79)

$$(32) \quad G(m) = \left(\frac{m}{2}\right)^{-T/2} \frac{I_{T/2}(m)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{T}{2} + \frac{1}{2}\right)}.$$

Differentiating each side of (32) k times, we find by (31) and (32)

$$\begin{aligned}
\frac{d^k}{dm^k} G(m) &= \int_{-1}^{+1} \bar{r}^k (1 - \bar{r}^2)^{T/2-1/2} \exp(m\bar{r}) d\bar{r} \\
(33) \quad &= \frac{2^{T/2}}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{T}{2} + \frac{1}{2}\right)} \frac{d^k}{dm^k} [m^{-T/2} I_{T/2}(m)].
\end{aligned}$$

Using the identity ([8], p. 79)

$$\frac{d}{dz} [z^{-\nu} I_\nu(z)] = z^{-\nu} I_{\nu+1}(z)$$

and changing the integration variable in (29) from p to m , we obtain finally

$$(34) \quad \tilde{E}[\bar{r}^k] = \frac{T}{2} \rho^{-T/2} \int_0^\infty m^{T/2-1} \exp\left(-\frac{m(1 + \rho^2)}{2\rho}\right) \frac{d^{k-1}}{dm^{k-1}} [m^{-T/2} I_{T/2+1}(m)] dm.$$

For $k = 1$, we have ([8], p. 386)

$$(35) \quad \tilde{E}[\bar{r}] = \frac{\rho}{1 + \frac{2}{T}}.$$

For $k = 2$, after some tedious calculation, we find

$$\begin{aligned}
\tilde{E}[\bar{r}^2] &= \frac{1}{T + 2} + \frac{\rho^2 T(T + 1)}{(T + 2)(T + 4)} \\
(36) \quad \tilde{\sigma}_{\bar{r}}^2 &= \frac{1}{T + 2} \left[1 - \frac{\rho^2 T(T - 2)}{(T + 2)(T + 4)} \right].
\end{aligned}$$

We note that $\lim_{T \rightarrow \infty} \tilde{E}(\bar{r}) = \rho$ and $\lim_{T \rightarrow \infty} \tilde{\sigma}_{\bar{r}}^2 = 0$, so that at least to the extent of approximation furnished by $\tilde{R}_{\rho}(\bar{r})$, \bar{r} is a consistent estimate of ρ .

The author wishes to express his gratitude to Dr. T. Koopmans, under whose kind direction this paper was written.

REFERENCES

- [1] T. KOOPMANS, "Serial correlation and quadratic forms in normal variables", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 14-33.
- [2] W. J. DIXON, "Further contributions to the problem of serial correlation", *Annals of Math. Stat.*, Vol. 15 (1944), pp. 119-144.
- [3] H. RUBIN, "On the distribution of the serial correlation coefficient", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 211-215.
- [4] R. L. ANDERSON, "Distribution of the serial correlation coefficient", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 1-13.
- [5] W. G. MADOW, "Note on the distribution of the serial correlation coefficient", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 308-310.
- [6] B. O. PEIRCE, *A Short Table of Integrals*, Ginn and Co., 1929.
- [7] G. A. CAMPBELL AND R. M. FOSTER, *Fourier Integrals for Practical Applications*, Bell Tel. Tech. Pub., 1942.
- [8] G. N. WATSON, *A Treatise on the Theory of Bessel Functions*, Second Rev. ed., Cambridge University Press, 1944.