ON THE DISTRIBUTION OF THE SERIAL CORRELATION COEFFICIENT

By Herman Rubin

Cowles Commission for Research in Economics, University of Chicago

The distribution of the serial correlation coefficient, in samples drawn from a parent distribution with zero serial correlation, has been studied by many authors. Anderson [1] obtained the exact distribution. Dixon [3] and Koopmans [4] have given approximate distributions, each attained by smoothing the characteristic values of the numerator of $\bar{r}$ in (1) below. Dixon smoothed the characteristic values in the generating function and obtained his results by comparing the moments of the exact distribution with those of the approximation, of which the first T are found to be exact. Koopmans smoothed the characteristic values in the exact distribution function. Here we evaluate Koopmans result and show that it is the same as Dixon's approximation. It thus appears that in this case it is immaterial whether the characteristic values are smoothed before or after inverting the characteristic function. We also add Tables comparing confidence limits for the exact distribution, for the approximation referred to, and for a normal approximation.

We define the serial correlation coefficient as

$$\bar{r} = \frac{\sum_{t=1}^{T} x_t x_{t+1}}{\sum_{t=1}^{T} x_t^2}, \quad x_{T+1} = x_1. \tag{1}$$

Then Koopmans obtains, if the true value ρ of $\bar{r}$ equals 0, and the x_t are normally and independently distributed with mean 0 and variance σ^2, the approximate distribution $T/2 - 2$.

$$\bar{h}(\bar{r}, T) = \frac{2^{\frac{1}{2}T}(\frac{1}{2}T - 1)}{\pi} \int_0^{\text{arc cos } \bar{r}} (\cos \alpha - \bar{r})^{\frac{1}{2}T-2} \sin \tfrac{1}{2}T\alpha \sin \alpha \, d\alpha. \tag{2}$$

Although in the distribution problem T is a positive integer, it is useful to consider the right-hand member of (2) as the definition of $\bar{h}(\bar{r} \cdot T)$ for those complex values of T for which it exists.

Let $R(T)$ denote the real part of T. If $R(T) > 2N + 2$, we obtain

$$\begin{aligned} \frac{d^r}{d\bar{r}^N} \bar{h}(\bar{r}, T) = {} & \frac{(-1)^N 2^{\frac{1}{2}T}(\frac{1}{2}T - 1)}{\pi} (\tfrac{1}{2}T - 2)(\tfrac{1}{2}T - 3) \cdots (\tfrac{1}{2}T - N - 1) \\ & \cdot \int_0^{\text{arc cos } \bar{r}} (\cos \alpha - \bar{r})^{\frac{1}{2}T-2-N} \sin \tfrac{1}{2}T\alpha \sin \alpha \, d\alpha. \end{aligned} \tag{3}$$

Now, according to [2], tables 41, 42.

$$\begin{aligned} & \int_0^{\pi/2} (\cos \alpha)^{\frac{1}{2}T-2-N} \sin \tfrac{1}{2}T\alpha \sin \alpha \, d\alpha \\ & \qquad = \frac{\frac{1}{2}T\pi}{2^{\frac{1}{2}T-N}} \frac{\Gamma(\frac{1}{2}T - N - 1)}{\Gamma(\frac{1}{2}(T - N + 1))\Gamma(\frac{1}{2}(1 - N))}. \end{aligned} \tag{4}$$

Deonote by $\bar{h}^{(N)}(0, T)$ the value of $\dfrac{d^N}{d\bar{r}^N}\,\bar{h}(\bar{r}, T)$ for $\bar{r} = 0$. Then for $R(T) > 2N + 2$,

$$h^{(N)}(0, T) = \frac{(-1)^N 2^N \Gamma(\frac{1}{2}T + 1)}{\Gamma(\frac{1}{2}(T - N + 1))\Gamma(\frac{1}{2}(1 - N))}. \tag{5}$$

$\bar{h}(\bar{r}, T)$ is analytic in $\bar{r}$ for $|\bar{r}| < 1$, $R(T) > 2$, and is analytic in T for $|\bar{r}| < 1$, $R(T) > 2$. It follows by Hartogs's theorem [5] that $\bar{h}(\bar{r}, T)$ is analytic in $\bar{r}$ and T for $|\bar{r}| < 1$, $R(T) > 2$. By analytic continuation we get that (5) holds for $R(T) > 2$. Consequently

(6) If N is odd, $\bar{h}^{(N)}(0, T) = 0$;

(7) if N is even,

$$\frac{\bar{h}^{(r)}(0, T)}{\bar{h}(0, T)} = \frac{2^N \Gamma(\frac{1}{2}T + \frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2}(T - N + 1))\Gamma(\frac{1}{2}(1 - N))}.$$

Let $N = 2P$, then

$$\begin{aligned}
&\frac{1}{(2P)!}\frac{\bar{h}^{(2P)}(0, T)}{\bar{h}(0, T)} \\
&= \frac{(-1)^P 2^{2P}}{(2P)!}\cdot\left(\frac{T-1}{2}\right)\cdot\left(\frac{T-3}{2}\right)\cdots\left(\frac{T-2P+1}{2}\right)\left(\frac{1\cdot 3\cdots(2P-1)}{2^P}\right) \\
&= \frac{(-1)^P}{(2P)!}\left(\frac{T-1}{2}\right)\left(\frac{T-3}{2}\right)\cdots\left(\frac{T-2P+1}{2}\right)\frac{(2P)!}{P!} \\
&= \frac{1}{P!}\left[\frac{d^P}{\{d(\bar{r}^2)\}^P}(1 - \bar{r}^2)^{\frac{1}{2}(T-1)}\right]_{\bar{r}=0}.
\end{aligned} \tag{8}$$

According to (5)

$$\bar{h}(0, T) = \frac{\Gamma(\frac{1}{2}T + 1)}{\Gamma(\frac{1}{2}T + \frac{1}{2})\Gamma(\frac{1}{2})} = \frac{1}{\int_{-1}^{1}(1 - \bar{r}^2)^{\frac{1}{2}(T-1)}\,d\bar{r}}. \tag{9}$$

Hence

$$\bar{h}(\bar{r}, T) = \frac{\Gamma(\frac{1}{2}T + 1)(1 - \bar{r}^2)^{\frac{1}{2}(T-1)}}{\Gamma(\frac{1}{2}T + \frac{1}{2})\Gamma(\frac{1}{2})}, \tag{10}$$

which is the same as Dixon's expression (3.22).

A more elementary proof by complete induction for integral values of T can be based on the recurrent differential equation (14) which is of interest in itself. To this end we shall write (2) in a different form which is easily obtained through partial integration.

$$\bar{h}(\bar{r}, T) = \frac{2^{\frac{1}{2}T}\cdot\frac{1}{2}T}{\pi}\int_0^{\arccos\bar{r}}(\cos\alpha - \bar{r})^{\frac{1}{2}T-1}\cos\tfrac{1}{2}T\alpha\,d\alpha. \tag{11}$$

Differentiating with respect to $\bar{r}$,

$$\bar{h}'(\bar{r}, T) = -\frac{\frac{1}{2}T(\frac{1}{2}T - 1)2^{\frac{1}{2}T}}{\pi} \int_0^{\text{arc cos } \bar{r}} (\cos\alpha - \bar{r})^{\frac{1}{2}T-2} \cos \tfrac{1}{2}\, T\alpha \, d\alpha.$$

$$= -\frac{\frac{1}{2}T(\frac{1}{2}T - 1)2^{\frac{1}{2}T}}{\pi} \int_0^{\text{arc cos } \bar{r}} (\cos\alpha - \bar{r})^{\frac{1}{2}T-2} (\cos \tfrac{1}{2}(T - 2)\, \alpha \cos\alpha - \sin \tfrac{1}{2}(T-2)\, \alpha \sin\alpha)\, d\alpha$$

(12)
$$= -\frac{\frac{1}{2}T(\frac{1}{2}T - 1)2^{\frac{1}{2}T}}{\pi} \int_0^{\text{arc cos } \bar{r}} (\cos\alpha - \bar{r})^{\frac{1}{2}T-1} \cos \tfrac{1}{2}(T-2)\alpha \, d\alpha$$
$$- \frac{\frac{1}{2}T(\frac{1}{2}T - 1)2^{\frac{1}{2}T}}{\pi} \int_0^{\text{arc cos } \bar{r}} (\cos\alpha - \bar{r})^{\frac{1}{2}T-2} \cos \tfrac{1}{2}(T-2)\alpha \, d\alpha$$
$$+ \frac{\frac{1}{2}T(\frac{1}{2}T - 1)2^{\frac{1}{2}T}}{\pi} \int_0^{\text{arc cos } \bar{r}} (\cos\alpha - \bar{r})^{\frac{1}{2}T-2} \cdot \sin \tfrac{1}{2}(T-2)\alpha \sin\alpha \, d\alpha$$

(13)
$$= -\frac{\frac{1}{2}T(\frac{1}{2}T - 1)2^{\frac{1}{2}T}\bar{r}}{\pi} \int_0^{\text{arc cos } \bar{r}} (\cos\alpha - \bar{r})^{\frac{1}{2}(T-2)-1} \cdot \cos \tfrac{1}{2}(T-2)\alpha \, d\alpha,$$

because the first and third terms in (12) cancel as may be shown by integrating by parts.

Hence (13) reduces to the recurrent differential equation

(14)
$$\bar{h}'(\bar{r}, T) = -2 \cdot \tfrac{1}{2}T\bar{r}\bar{h}(\bar{r}, T - 2).$$

Let us now assume that

(15)
$$\bar{h}(\bar{r}, T - 2) = \frac{\Gamma(\frac{1}{2}T)}{\Gamma(\frac{1}{2}T - \frac{1}{2})\Gamma(\frac{1}{2})} (1 - \bar{r}^2)^{\frac{1}{2}(T-3)}.$$

Then (14) becomes

(16)
$$\bar{h}(\bar{r}, T) = -2\bar{r}\cdot\tfrac{1}{2}T \frac{\frac{1}{2}(T-1)}{\frac{1}{2}T - \frac{1}{2}} \frac{\Gamma(\frac{1}{2}T)}{\Gamma(\frac{1}{2}T - \frac{1}{2})\Gamma(\frac{1}{2})} (1 - \bar{r}^2)^{\frac{1}{2}T-3}$$
$$= -2\bar{r}\cdot\tfrac{1}{2}(T-1) \frac{\Gamma(\frac{1}{2}T + 1)}{\Gamma(\frac{1}{2}T + \frac{1}{2})\Gamma(\frac{1}{2})} (1 - \bar{r}^2)^{\frac{1}{2}(T-1)-1}.$$

Integrating, one obtains

(17)
$$\bar{h}(\bar{r}, T) = \frac{\Gamma(\frac{1}{2}T + 1)}{\Gamma(\frac{1}{2}T + \frac{1}{2})\Gamma(\frac{1}{2})} (1 - \bar{r}^2)^{\frac{1}{2}(-2)}.$$

No constant of integration occurs because (17) agrees with (5) for $\bar{r} = 0$ and $N = 0$.

It remains to prove the validity of (17) for the initial values $T = 3$ and $T = 4$. If $T = 4$

$$\begin{aligned}\bar{h}(\bar{r}, 4) &= \frac{4}{\pi}\int_0^{\operatorname{arc\,cos}\bar{r}} \sin 2\alpha \sin\alpha \, d\alpha \\ &= \frac{8}{3\pi}\sin^3\alpha \Big|_0^{\operatorname{arc\,cos}\bar{r}} = \frac{8}{3\pi}(1-\bar{r}^2)^{\frac{3}{2}} = \frac{\Gamma(3)}{\Gamma(\frac{5}{2})\Gamma(\frac{1}{2})}(1-\bar{r}^2)^{\frac{1}{2}(4-1)}.\end{aligned} \tag{18}$$

For $T = 3$,

$$\bar{h}(\bar{r}, 3) = \frac{\frac{1}{2}2^{\frac{1}{2}}}{\pi}\int_0^{\operatorname{arc\,cos}\bar{r}} \frac{\sin\frac{3}{2}\alpha \sin\alpha}{\sqrt{\cos\alpha - \bar{r}}}\, d\alpha. \tag{19}$$

Substitute $\cos\alpha = \bar{r} + (1 - \bar{r})\sin^2\theta$. We get

$$\begin{aligned}\bar{h}(\bar{r}, 3) &= \frac{2(1-\bar{r})}{\pi}\int_0^{\frac{1}{2}\pi} \{(1 + 2\bar{r})\cos^2\theta + 2(1-\bar{r})\sin^2\theta\cos^2\theta\}\, d\theta \\ &= \tfrac{3}{4}(1-\bar{r}^2) = \frac{\Gamma(\frac{5}{2})}{\Gamma(2)\Gamma(\frac{1}{2})}(1-\bar{r}^2)^{\frac{1}{2}(3-1)}\end{aligned} \tag{20}$$

which completes the proof.

A short table of confidence limits is included, corresponding to the 5% and 1% significance levels, comparing the exact distribution given by Anderson [1] (the values in parentheses being graphically interpolated by him), the distribution (10), and the normal curve with the same mean and standard deviation.

Confidence limits for $\bar{r}$

T	5%			1%		
	Exact	(10)	Normal	Exact	(10)	Normal
3	.854	.729	.736	.970	.882	1.040
4	.713	.669	.672	.898	.833	.950
5	.622	.621	.622	.823	.789	.879
6	.570	.582	.582	.762	.750	.823
7	.545	.549	.548	.714	.715	.775
8	(.521)	.521	.520	(.682)	.685	.736
9	.498	.497	.496	.656	.658	.701
10	(.477)	.476	.475	(.633)	.634	.672
11	.457	.458	.456	.612	.612	.645
15	.400	.400	.399	.543	.543	.564
20	(.351)	.352	.351	(.480)	.482	.496
25	.317	.317	.317	.437	[illegible]37	.448
30	(.291)	.291	.291	(.404)	.403	.411
35	(.271)	.271	.270	(.377)	.376	.382
40	(.255)	.254	.254	(.355)	.354	.359
45	.240	.240	.240	.335	.335	.339

It is thus seen that the distribution (10) provides satisfactory significance levels for $T \geq 9$ whereas the normal approximation provides satisfactory 5% significance levels for the same range. The normal approximation appears to be unsatisfactory, however, at the 1% significance level even for T as high as 45. The normal approximation here used is not the same as that used by Anderson ([1], p. 53), which assumes $\dfrac{\sqrt{T}\,\bar{r}}{\sqrt{1 + 2\bar{r}^2}}$ to be normally distributed.

The following table shows a comparison between a few more confidence limits of the Type II curve (10) and the normal curve with same first two moments for a few values of T.

Confidence limits for $\bar{r}$

T	5%		4%		3%		2%		1%	
	(10)	Normal	(10)	Normal	(10)	Normal	(10)	Normal	(10)	Normal
15	.400	.399	.423	.425	.452	.456	.488	.498	.543	.564
20	.352	.351	.373	.373	.398	.401	.431	.438	.482	.496
25	.317	.317	.336	.337	.360	.362	.390	.395	.437	.448

REFERENCES

[1] R. L. Anderson, *Serial Correlation in the Analysis of Time Series*, unpublished thesis, Iowa State College, 1941.

[2] D. Bierens de Haan, *Nouvelles Tables d'Integrales Definies*, Leyden, 1867.

[3] T. Koopmans, "Serial Correlation and Quadratic Forms in Normal Variables", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 14–33.

[4] W. J. Dixon, "Further contributions to the problem of serial correlation", *Annals of Math. Stat.*, Vol. 14 (1944).

[5] W. F. Osgood, *Lehrbuch der Functionentheorie*, Vol. 2. Part 2, Leipzig, 1907.