Anim. Behav., 1996, **51,** 1175–1178

COMMENTARIES

Comments on predictability and overdifferencing in vigilance sequences

MATTHEW KRAMER & WILLIAM R. BELL
U.S. Bureau of the Census, Washington D.C., U.S.A.

(Received 5 December 1994; initial acceptance 14 February 1995; final acceptance 26 May 1995; MS. number: sc-1053)

The degree of predictability in behaviour sequences is a central ethological question in many studies. For example, individuals, while routinely scanning for predators or locations of conspecifics, may be at a disadvantage if their scans can be predicted. Two empirical issues arise in regard to the predictability of these vigilance sequences, here defined as time series consisting of durations of sequential inter-scan intervals (ISIs). First, to what degree are ISI durations predictable from observations of past ISI durations? Second, does knowledge of the time spent in the current interval (time since the previous scan) provide information useful for predicting how much longer the current interval will last? Desportes et al. (1989) refer to unpredictability of ISI durations based on past ISI durations as sequential randomness, and unpredictability of the remaining length of an ISI based on knowledge of the time spent in the current interval as instantaneous randomness. The first issue relates to time series (autocorrelation) properties of ISI durations, the second to the probability distribution of ISI durations. These two issues are essentially distinct, both conceptually and statistically. In this paper we discuss the first issue in reference to a recent article of Roberts (1994).

In his article, Roberts (1994) examined the first issue by using time series methods described by Box & Jenkins (1976) to analyse sequences of ISI durations in preening sanderlings, *Calidris alba*. The first step in this approach is to examine estimates of autocorrelations. The lag k autocorrelation is $\mathrm{corr}(Y_{t+k}, Y_t)$, where Y_t denotes the time series of ISI durations, indexed by t. Having examined estimated autocorrelations for 13 time series of ISI durations, Roberts (1994, page 584) reported that, 'sequences of ISI durations appear to be essentially random', i.e. uncorrelated. (Figure 2a of Roberts (1994) gives the autocorrelation function for a representative sequence of ISI durations.) The time series analysis should have stopped at this point, because if there is no evidence of autocorrelation in the data there is no need for a time series model. In fact, the goal in time series modelling is to find a model that transforms the original time series into an uncorrelated series with mean zero (called a white noise series). The model that accomplishes this for Y_t is the simple mean model:

$$Y_t = \mu + e_t \quad (1)$$

where μ is the mean and e_t is the white noise series.

Roberts (1994), however, differenced his time series and performed various analyses on the differenced series, $W_t = Y_t - Y_{t-1}$. Roberts (1994) stated that differencing was performed to 'remove the effects of long term trends in the sequences' (page 581). The nature of such trends is not specified. Although differencing is a useful tool in time series modelling, it is a powerful transformation that should not be used indiscriminately. Box & Jenkins (1976, Section 6.2.1), suggested differencing only for time series whose sample autocorrelations fail to die out rapidly with increasing lag. Such behaviour of autocorrelations reflects non-stationarity in the series that can arise either from stochastic trends that require differencing, or from deterministic polynomial trends (e.g. the linear trend $\alpha + \beta t$), which can also be rendered stationary by differencing. A time series is stationary if its mean $\mathrm{E}(Y_t)$ and variance $\mathrm{var}(Y_t)$ are constant over time t, and if its theoretical autocorrelations $\mathrm{corr}(Y_t, Y_{t+k})$ depend only on lag k and not on t. Differencing is clearly

Correspondence: M. Kramer, Statistical Research Division, Room 3000-4, U.S. Bureau of the Census, Washington D.C. 20233-9100, U.S.A.

0003–3472/96/051175+04 $18.00/0

inappropriate for time series that give no evidence of autocorrelation, such as the time series whose autocorrelation function is given in Figure 2a of Roberts (1994).

Inducing autocorrelation by unnecessary differencing will not extract additional information from a time series. In fact, the results obtained by Roberts from analyses of the differenced time series of ISI durations are purely artefacts of the unnecessary differencing, and admit no meaningful interpretation. We demonstrate this for results that Roberts (1994) obtained from fitting autoregressive (AR) models to his differenced ISI durations, by showing that his results are very close to what would be expected from fitting AR models to a differenced white noise series. Analogous analyses could be presented to show that the other results obtained by Roberts (1994) are also what would be expected from a differenced white noise series.

First, note that if Y_t follows the mean model (1), then in differencing Y_t to $W_t = Y_t - Y_{t-1}$, the mean μ drops out and the model followed by the differenced series W_t is

$$W_t = e_t - e_{t-1}. \qquad (2)$$

Because successive observations W_t and W_{t+1} involve an overlap of e_t, the W_t series will be autocorrelated at lag one. That is, $\mathrm{cov}(W_{t+1}, W_t) = \mathrm{cov}(e_{t+1} - e_t, e_t - e_{t-1}) = -\mathrm{cov}(e_t, e_t) = -\mathrm{var}(e_t)$. Observations of W_t more than one time point apart involve no overlap of e_t, and hence are uncorrelated.

Model (2) is actually a special case of the first-order moving-average (MA(1)) model discussed by Box & Jenkins (1976, Section 3.3). The general pattern of autocorrelations for MA(1) models are a non-zero value at lag 1, and zero at lags 2, 3, ... For the particular MA(1) model given in (2), the theoretical lag 1 autocorrelation, $\rho_1 = \mathrm{corr}(W_{t+1}, W_t)$, is $-\frac{1}{2}$, and $\rho_2 = \rho_3 = \cdots = 0$ for lags 2, 3, ... (Box & Jenkins 1976, equation 3.3.4). Estimated autocorrelations from data following model (2) would be expected to exhibit a similar pattern. In fact, the estimated autocorrelations for the representative differenced series given in Figure 2b of Roberts (1994) show (within estimation standard errors) just this pattern. For this differenced series, the observed lag 1 autocorrelation is approximately -0.5, and autocorrelations are not different from zero at all other lags.

Roberts (1994, Table I), however, presented results from fitting AR models to the differenced duration series. AR models are inappropriate for these data, since the correct model for the differenced series, given that the original series is uncorrelated, is the MA(1) model given in equation (2). This particular model will be poorly approximated by AR models. Nevertheless, we can ask what results would be expected if AR models are fitted to a series W_t that actually follows the MA(1) model (2).

The behaviour of autoregressive parameter estimates when fitting models to a series that follows equation (2) can be understood by considering the Yule–Walker equations (Box & Jenkins 1976, equation 3.2.6). These equations relate parameters of an AR model of specified order p to the autocorrelations at lags $1, \ldots, p$. If theoretical autocorrelations $\rho_1, \ldots, \rho_p$ are known and the true model is AR(p), then solving the corresponding Yule–Walker equations yields the AR parameters $\phi_1, \ldots, \phi_p$. If the true model is not AR(p), then the solution to the Yule–Walker equations indicates what the parameter estimates obtained from fitting an AR(p) model are actually estimating. For the first order autoregressive model (AR(1)) the lone Yule–Walker equation is $\phi_1 = \rho_1$, where ϕ_1 is the AR(1) parameter. Thus, for the AR(1) model, the estimate of ϕ_1 is estimating the lag 1 sample autocorrelation (regardless of whether the true model is actually AR(1)). For a differenced white noise series, we would thus expect that fitting AR(1) models would produce estimates of ϕ_1 approximately equal to $\rho_1 = -\frac{1}{2}$ (within estimation error).

The behaviour of autoregressive parameter estimates when fitting a higher order AR(p) model ($p>1$) to a differenced white noise series can be understood by substituting $\rho_1 = -\frac{1}{2}$ and $\rho_k = 0$ for $k=2, \ldots, p$ in the Yule–Walker equations, and then solving the resulting equations for $\phi_1, \ldots, \phi_p$. For example, consider the AR(3) model. The Yule–Walker equations for the AR(3) are

$$\rho_1 = \phi_1 + \phi_2\rho_1 + \phi_3\rho_2$$

$$\rho_2 = \phi_1\rho_1 + \phi_2 + \phi_3\rho_1$$

$$\rho_3 = \phi_1\rho_2 + \phi_2\rho_1 + \phi_3,$$

which, in matrix notation, are

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix}.$$

Table I. Comparison of the autoregressive parameters given in the first six sequences of Roberts (1994) (R) with theoretical values (T) from fitting the same order autoregressive model to an overdifferenced white noise series

Sequence	AR (1)	AR (2)	AR (3)	AR (4)	AR (5)	AR (6)	AR (7)
1R	−0.81	−0.54	−0.26				
1T	−0.75	−0.50	−0.25				
2R	−0.63	−0.61	−0.50	−0.38	−0.39		
2T	−0.83	−0.67	−0.50	−0.33	−0.17		
3R	−0.93	−0.89	−0.66	−0.69	−0.60	−0.49	−0.35
3T	−0.88	−0.75	−0.63	−0.50	−0.38	−0.25	−0.13
4R	−0.73	−0.42	−0.43	−0.27			
4T	−0.80	−0.60	−0.40	−0.20			
5R	−0.97	−0.76	−0.71	−0.59	−0.66	−0.50	
5T	−0.86	−0.71	−0.57	−0.43	−0.29	−0.14	
6R	−0.80	−0.50					
6T	−0.67	−0.33					

Substituting in $\rho_1 = -\frac{1}{2}$ and $\rho_2 = \rho_3 = 0$ gives

$$\begin{pmatrix} -\frac{1}{2} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix}.$$

The solution for ϕ_1, ϕ_2 and ϕ_3 is

$$\begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & 1 \end{pmatrix}^{-1} \begin{pmatrix} -\frac{1}{2} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{3}{4} \\ -\frac{1}{2} \\ -\frac{1}{4} \end{pmatrix}.$$

The interpretation of this result is that if an AR(3) model is (incorrectly) fitted to a differenced white noise series as in equation (2), the AR(3) parameter estimates will be estimating -0.75, -0.50 and -0.25. For comparison, the AR(3) parameter estimates obtained by Roberts (1994, Table I) for his first series are -0.81, -0.54 and -0.26.

In general, when an AR(p) model is fitted to an unnecessarily differenced white noise series (2), the values being estimated by the autoregressive parameter estimates $\hat{\phi}_1, \ldots, \hat{\phi}_p$ are $-p/(p+1), \ldots, -1/(p+1)$. For series of the length Roberts (1994) considered, the estimates $\hat{\phi}_1, \ldots, \hat{\phi}_p$ are subject to substantial error. Given this, note how closely the values given in Table I in Roberts (1994) agree with the pattern $-p/(p+1), \ldots, -1/(p+1)$ (see Table I). The AR fitting results are thus completely consistent with the hypothesis that the original series are uncorrelated.

A central point of Roberts' (1994) article seems to rest on the misconception that differencing the series of ISI durations accomplishes something since the differenced series exhibit negative autocorrelation at lag one, and hence some predictability that is not present in the original series (which appear uncorrelated). In fact, it should not be surprising that the differenced series is somewhat predictable from its past. Having taken N observations, the first future value of the differenced series will be $W_{N+1} = Y_{N+1} - Y_N$. Since Y_N is the last observed value, this part of W_{N+1} is known, and hence perfectly predictable. The other part of W_{N+1}, Y_{N+1}, is 'unpredictable' in that it is uncorrelated with the data through time N. Future values of the differenced series beyond time $N+1$ are uncorrelated with available data, since they are determined exclusively by future values Y_t for $t>N$.

Actually, some care needs to be taken in regard to interpretations of 'predictability' or 'unpredictability'. The fact that the original time series of ISI durations appear to be uncorrelated over time does not necessarily mean that they are unpredictable, or that past data are entirely irrelevant to their prediction. For the mean model (1) the best prediction of any future value of Y_t is the mean μ, which can be estimated by the sample mean $\bar{Y}$. It is in estimating μ by $\bar{Y}$ that past data are used to form the prediction. The variance of the error in the prediction of Y_t for the mean model (1) for any $t>N$, where N is the number of observations, can be calculated using standard regression prediction results (e.g. Draper & Smith 1981, equation 1.4.11). This is given as

$$\mathrm{var}(Y_t - \bar{Y}) = \sigma^2(1 + 1/N), \tag{3}$$

where $\sigma^2 = \mathrm{var}(Y_t)$. The magnitude of this variance thus depends on σ^2 and on the number of observations N. Unless N is very small, say $N<10$, the dominant factor will be σ^2. For non-negative data, if σ is considerably smaller than μ, then the coefficient of variation (CV) for predicted Y_t is small and ISI durations will be fairly predictable (using the estimated mean). If the CV is large, then ISI durations will be fairly unpredictable. Roberts (1994) reported that his 13 ISI duration time series had means and corresponding standard errors ($\bar{Y} \pm \hat{\sigma}/\sqrt{N}$) ranging from 1.51 ± 0.12 s to 3.56 ± 0.32 s. CVs calculated by dividing equation (3) by $\bar{Y}$ cannot be obtained from this information without knowing the number of observations N for each series. It is reported only that N is at least 40 and is a maximum of 69 for the 13 time series. Taking the two examples cited, and the minimum and maximum values of N, the CV for the first case would be between 0.51 $(=(0.12\sqrt{40}\sqrt{1+1/40})/1.51)$ for $N=40$ and 0.66 for $N=69$, while for the second case it would be between 0.58 ($N=40$) and 0.75 ($N=69$). Since the CVs in both cases are large, the ISI durations for these two cases are indeed largely unpredictable from past duration data.

Conversely, just because a time series is autocorrelated, even strongly autocorrelated, does not mean it is predictable from past data. Using an appropriate time series model for the data, if the prediction standard errors (Box & Jenkins 1976, Section 5.1) are large relative to the predictions, then the series is essentially unpredictable, even though autocorrelated.

In his discussion, Roberts (1994) postulated a behavioural mechanism that could lead to negative autocorrelation in the differenced ISI durations. We have shown, however, that the same numerical (statistical) results would be expected from unnecessarily differencing a white noise series. Thus, the data offer no evidence in support of Roberts' (1994) postulated mechanism. This mechanism involves the animal internally tracking past differenced ISI durations and using these in determining its future scanning behaviour, through a mechanism that depends on a variance, several autoregressive parameters and an internal random number generator. The data, in fact, support the much simpler explanation that no internal tracking of past behaviour is performed, and the lengths of ISI durations are random, independently generated from some probability distribution that might be characterized by its mean and variance.

We thank two referees for constructive criticism, especially for helping to emphasize certain points.

REFERENCES

Box, G. E. P. & Jenkins, G. M. 1976. *Time Series Analysis: Forecasting and Control.* Revised edn. San Francisco: Holden-Day.

Desportes, J.-P., Metcalfe, N. B., Cézilly, F., Lauvergeon, G. & Kervella, C. 1989. Tests of the sequential randomness of vigilant behaviour using spectral analysis. *Anim. Behav.*, **38,** 771–777.

Draper, N. R. & Smith, H. 1981. *Applied Regression Analysis.* 2nd edn. New York: John Wiley.

Roberts, G. 1994. When to scan: an analysis of predictability in vigilance sequences using autoregression models. *Anim. Behav.*, **48,** 579–585.