FIRST-ORDER INTEGER-VALUED AUTOREGRESSIVE (INAR(1)) PROCESS

By M. A. Al-Osh and A. A. Alzaid

King Saud University

Abstract. A simple model for a stationary sequence of integer-valued random variables with lag-one dependence is given and is referred to as the integer-valued autoregressive of order one (INAR(1)) process. The model is suitable for counting processes in which an element of the process at time t can be either the survival of an element of the process at time $t - 1$ or the outcome of an innovation process. The correlation structure and the distributional properties of the INAR(1) model are similar to those of the continuous-valued AR(1) process. Several methods for estimating the parameters of the model are discussed, and the results of a simulation study for these estimation methods are presented.

Keywords. INAR(1) model; convolution; conditional least squares; conditional maximum likelihood.

1. INTRODUCTION

Statistical data which are expressed in terms of counts taken sequentially in time, and which are correlated, arise in many settings. Examples of this process are the number of patients in a hospital at a specific point of time, or the number of persons in a queue waiting for service at a certain moment. In each of these examples an element of the process at time t can be either the survival of an element of the process at previous times, or an arrival (innovation) sequence which has a certain discrete distribution. The purpose of this paper is to model such a process in a manner analogous to the continuous-valued autoregressive-moving average (ARMA) process, in which the innovation sequence is often assumed, for statistical convenience, to have normal distribution.

In the past several years, various models have been proposed for stationary discrete time series. Jacobs and Lewis (1978a, b, 1983) introduced what they have called discrete autoregressive-moving average (DARMA) models, which were obtained by a probabilistic mixture of a sequence of independent identically distributed (i.i.d.) discrete random variables. As an example of this process, consider the DARMA(1, 1) model which is defined for $t = 1, 2, \ldots$ as follows

$$X_t = \begin{cases} Y_t & \text{with probability } \beta \\ A_{t-1} & \text{with probability } (1 - \beta) \end{cases}$$

with

$$A_t = \begin{cases} A_{t-1} & \text{with probability } \rho \\ Y_t & \text{with probability } (1 - \rho) \end{cases}$$

where $\{Y_t\}$ is a sequence of i.i.d. discrete random variables with common distribution π, and A_0 has the distribution π independent of $\{Y_t; t \geq 1\}$.

In the DARMA(1, 1) process, A_{t-1} contains all information available about the past, as do X_{t-1} and the random shock a_{t-1} in the ARMA(1, 1). However, in the discrete model there is randomization between A_{t-1} and Y_t, and if Y_t is chosen, then the memory of the process before time t is lost forever. This is in contrast to the memory of the continuous-valued time series which decays exponentially from the first-order autocorrelation (see, e.g., Box and Jenkins, 1976 Chap. 3).

The Jacobs–Lewis approach can be viewed as analogous to the Box–Jenkins approach in which the linear combination of the continuous-valued case is replaced with a probabilistic mixture. As a result of this mixing, a realization of the process will generally, as the authors remarked, contain many runs of a single value.

This paper deals with modelling integer-valued time series in a manner analogous to the standard time series models. Since the general model cannot be presented in great detail here, the paper is restricted to the integer-valued autoregressive of order 1, INAR(1), model. Construction of the INAR(1) model and its relation to a closely related model by Ali Khan (1986) are presented in the following section. Section 3 deals with the autocorrelation and distributional properties of the model. Section 4 presents an estimation of the parameters of the model when the innovation sequence has a Poisson distribution, along with the results of a simulation study concerning the estimation methods. Finally, some concluding remarks are made about the extension of the model and its applicability.

2. THE INTEGER-VALUED AUTOREGRESSIVE OF ORDER 1, INAR(1), MODEL

Before introducing the INAR(1) model, we present the definition of the '∘' operator which is due to Steutal and Van Harn (1979), see also Van Harn (1978, p. 85).

Let X be a non-negative integer-valued random variable; then for any $\alpha \in [0, 1]$ the operator '∘' is defined by

$$\alpha \circ X = \sum_{i=1}^{X} Y_i$$

where Y_i is a sequence of i.i.d. random variables, independent of X, such that

$$P_r(Y_i = 1) = 1 - P_r(Y_i = 0) = \alpha.$$

From the definition of the '∘' operator it is clear that $0 \circ X = 0$, $1 \circ X = X$, $E(\alpha \circ X) = \alpha E(X)$ and for any $\beta \in [0, 1]$, $\beta \circ (\alpha \circ X) \stackrel{d}{=} (\beta\alpha) \circ X$. Now we define the INAR(1) process $\{X_t: t = 0, \pm 1, \pm 2, \ldots\}$ by

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t \tag{2.1}$$

where $\alpha \in [0, 1]$ and ε_t is a sequence of uncorrelated non-negative integer-valued random variables having mean μ and finite variance σ^2.

The INAR(1) model defined by (2.1) simply states that the components of the process at time t, X_t, are (i) the survivors of the elements of the process at time $t - 1$, X_{t-1}, each with probability of survival α and (ii) elements which entered the system in the interval $(t - 1, t]$ as innovation term (ε_t).

As an illustration of how the sample path of the INAR(1) process differs from that of the standard AR(1) process, a simulated sample path for each process was generated. For comparison of the two sample paths, the two simulated processes are taken to have the same mean and the same variance. Figure 1 presents the simulated sample path based on a sample of size 100 observations generated from model (2.1), in which Y_i is taken to be Bernoulli random variables with $\alpha = 0.5$ and ε_t is taken to be Poisson with parameter $\lambda = 1.0$. Figure 2 shows the corresponding sample path of the standard AR(1) model, in which ε_t is taken to be normally distributed with mean λ and variance $\lambda(1 + \alpha)$.

It can be seen from the comparison of Figures 1 and 2 that whereas realizations of the INAR(1) process take non-negative integer values, realizations of the AR(1) process are real values some of which are negative. As a consequence of this observation and the stationarity requirements for the two processes, it can be seen from Figure 1 that realizations of the INAR(1) process have many runs at/near the mean (2). However, such a pattern of runs does not exist in Figure 2.
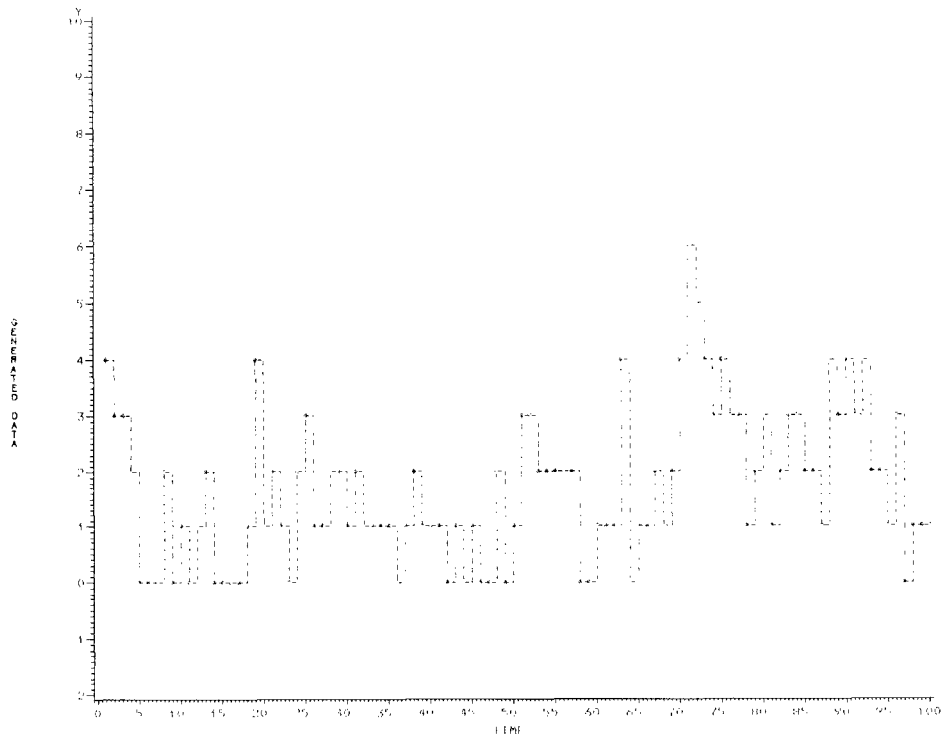

FIGURE 1.   Integer valued time series generated from Bernoulli (0.5) and Poisson (1.0).
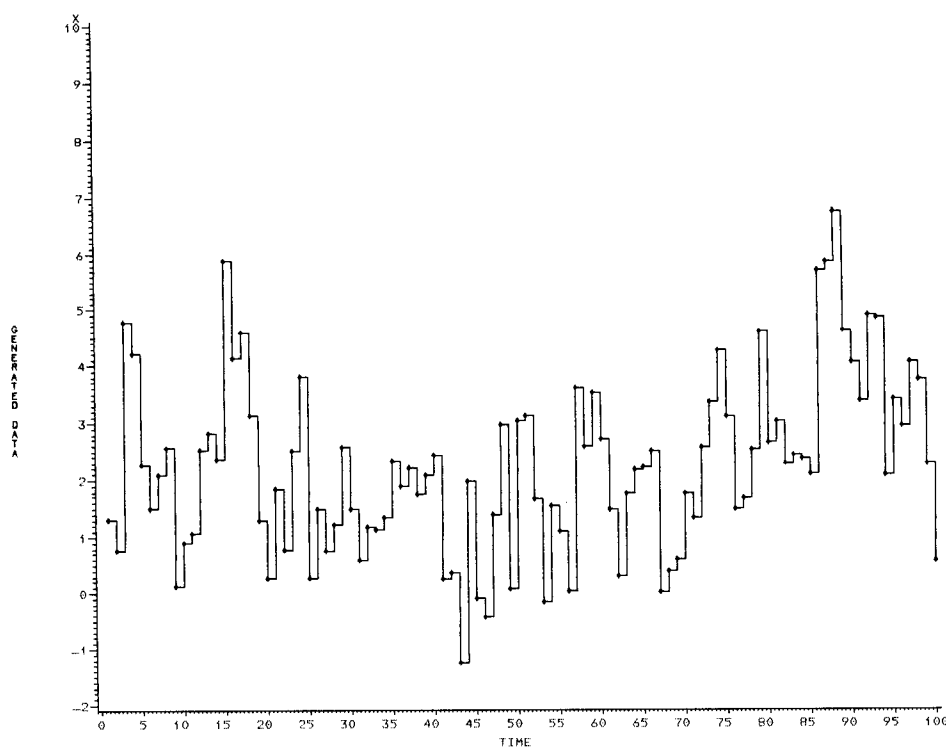
FIGURÈ 2.   Continuous time series generated from normal (2, 2).

The marginal distribution of model (2.1) can be expressed in terms of the innovation sequence $\{\varepsilon_t\}$ as

$$X_t \stackrel{\mathrm{d}}{=} \sum_{j=0}^{\infty} \alpha^j \circ \varepsilon_{t-j}. \tag{2.2}$$

For $\alpha \in (0, 1)$, it can be seen from (2.2) that the dependence of X_t on the sequence $\{\varepsilon_t\}$ decays exponentially with the time lag of this sequence.

We may also point out that

$$(X_t, X_{t-k}) \stackrel{\mathrm{d}}{=} \left(\alpha^k \circ X_{t-k} + \sum_{i=0}^{k-1} \alpha^i \circ \varepsilon_{t-i}, X_{t-k} \right).$$

As an example of the applicability of the INAR(1) model in other disciplines we consider the relation between this model and a recently proposed model by Ali Khan (1986) in terms of dam theory which has the following form

$$X_t = X_{t-1} + Z_{t-1} - U_t \qquad t = 1, 2, \ldots \tag{2.3}$$

where X_t, Z_t and U_t represent the content of the dam at the end of time t, the input during the time interval $[t, t + 1)$ and the output at the end of time t, respectively, so that given $X_{t-1} + Z_{t-1}$, U_t has a binomial distribution with

index parameter $(X_{t-1} + Z_{t-1})$ and scale parameter β. In terms of the operator '$\circ$', the model (2.3) can be written as

$$X_t = \alpha \circ (X_{t-1} + Z_{t-1}).$$

Equivalently

$$X_t \overset{d}{=} \alpha \circ X_{t-1} + \varepsilon_t$$

where $\alpha = 1 - \beta$ and $\varepsilon_t = \alpha \circ Z_{t-1}$. The last representation of X_t implies that Ali Khan's model can be viewed as an INAR(1) process in which ε_t is independent of X_{t-1}.

3. CORRELATION AND DISTRIBUTIONAL PROPERTIES OF THE INAR(1) MODEL

The mean and the variance of the process $\{X_t\}$ as defined in (2.1) are simply

$$E(X_t) = \alpha E(X_{t-1}) + \mu = \alpha^t E(X_0) + \mu \sum_{j=0}^{t-1} \alpha^j \qquad (3.1)$$

and

$$\text{var}(X_t) = \alpha^2 \text{var}(X_{t-1}) + \alpha(1 - \alpha)E(X_{t-1}) + \sigma^2$$

$$= \alpha^{2t} \text{var}(X_0) + (1 - \alpha) \sum_{j=1}^{t} \alpha^{2j-1} E(X_{t-j}) + \sigma^2 \sum_{j=1}^{t} \alpha^{2(j-1)}. \qquad (3.2)$$

It can be seen from (3.1) and (3.2) that second-order stationarity requires that the initial value of the process, X_0, has $E(X_0) = \mu/(1 - \alpha)$ and $\text{var}(X_0) = (\alpha\mu + \sigma^2)/(1 - \alpha^2)$. For any non-negative integer k, the covariance at lag k, $\gamma(k)$, is

$$\gamma(k) = \text{cov}(X_{t-k}, X_t) = \text{cov}(X_{t-k}, \alpha^k \circ X_{t-k}) + \text{cov}\left(X_{t-k}, \sum_{j=0}^{k-1} \alpha^j \circ \varepsilon_{t-j}\right)$$

$$= \alpha^k \text{var}(X_{t-k}) + \sum_{j=0}^{k-1} \alpha^j \text{cov}(X_{t-k}, \varepsilon_{t-j}) = \alpha^k \gamma(0). \qquad (3.3)$$

Equation (3.3) shows that the autocorrelation function, $\rho(k)$, decays exponentially with lag k and has the same form as the Yule–Walker equation in the AR(1). However, unlike the autocorrelation function for the standard AR(1), $\rho(k)$ is always positive for the INAR(1) model.

The derivation of $\gamma(k)$ by using the form (2.2) of the model is straightforward, after noting that

(i) $\text{cov}(\alpha^k \circ \varepsilon_{t-k}, \alpha^s \circ \varepsilon_{t-l}) = 0$ for $l \neq k$ and any positive integer s and

(ii) $\alpha^{j+k} \circ \varepsilon_{t-j} \overset{d}{=} \alpha^k \circ \alpha^j \circ \varepsilon_{t-j}$.

Regarding the distribution properties of the INAR(1) process, one remarks that this process has distributional properties similar to those of the AR(1). For example, when $X_t \overset{d}{=} X_{t-1}$, and X_{t-1} is independent of ε_t in (2.1), the validity of (2.1) for all $\alpha \in (0, 1)$ implies that X_t has a discrete self-decomposable distribution (see e.g. Stuetal and Van Harn, 1979). Having mentioned this property, other properties for X_t then follow, such as unimodality and the fact that the distribution of X_t is uniquely determined by the distribution of ε_t. As a result of the last

property, it follows that X_t has a Poisson distribution if and only if ε_t has a Poisson distribution. Consequently, the role which the distribution of ε_t plays in determining the distribution of X_t in the INAR(1) process is similar to the role played by the normal distribution of the errors in the AR(1) process. For other properties of the discrete self-decomposable distribution we refer to Stuetal and Van Harn (1979) and Van Harn (1978). We should mention here that the self-decomposability of the first autoregressive process was considered by Gaver and Lewis (1980) and Bondensson (1981).

4. ESTIMATION

The estimation problem connected with the INAR(1) process is more complicated than that of the AR(1) process. The complication arises from the fact that the conditional distribution of X_t given X_{t-1} in the INAR(1) process is the convolution of the distribution of ε_t and that of a binomial with scale parameter α and index parameter X_{t-1}. Throughout the present section we will assume that the sequence $\{\varepsilon_t\}$ has a Poisson distribution with parameter λ. In what follows we will present several methods for estimating the parameters α and λ.

4.1. *The Yule–Walker estimators*

The simplest way to get an estimator for α is to replace $\gamma(k)$ with the sample autocovariance function in the Yule–Walker equation (3.3) and solve for α to obtain

$$\hat{\alpha} = \frac{\sum\limits_{t=0}^{n-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum\limits_{t=0}^{n} (x_t - \bar{x})^2} \tag{4.1}$$

where $\bar{x}$ is the sample mean.

Now an estimate for λ can be obtained by calculating $\hat{\varepsilon}_t = x_t - \hat{\alpha}x_{t-1}$ for $t = 1, 2, \ldots, n$. Since it has been assumed that X_t has a Poisson distribution with parameter λ, then a reasonable estimator for λ is

$$\hat{\lambda} = \sum\limits_{t=1}^{n} \hat{\varepsilon}_t / n. \tag{4.2}$$

4.2. *The conditional least squares estimators*

Unlike the standard AR(1) model, one finds for the INAR(1) process that given X_{t-1} and ε_t, X_t is still a random variable. The conditional mean of X_t given X_{t-1} is given by

$$E(X_t/X_{t-1}) = \alpha X_{t-1} + \lambda \equiv g(\theta, X_{t-1}) \quad \text{say}$$

where $\theta = (\alpha, \lambda')$ is the set of parameters to be estimated. The estimation procedure which we are going to apply was developed by Klimko and Nelson (1978)

and it is based on minimization of the sum of squared deviations about the conditional expectation. Thus the conditional least squares (CLS) estimators for α and λ are those values which minimize

$$Q_n(\theta) = \sum_{t=1}^{n} [(X_t - g(\theta, X_{t-1})]^2 \tag{4.3}$$

with respect to θ. In evaluating the derivatives of (4.3) with respect to α and λ one obtains

$$\hat{\alpha} = \frac{\sum x_t x_{t-1} - (\sum x_t \sum x_{t-1})/n}{\sum x_{t-1}^2 - (\sum x_{t-1})^2/n} \tag{4.4}$$

and

$$\hat{\lambda} = \frac{1}{n} (\sum x_t - \hat{\alpha} \sum x_{t-1}) \tag{4.5}$$

where all summations are over t, which takes values from 1 to n.

Since one would expect, for stationary processes, that estimates for the mean and the variance calculated from the first n observations would be very close to those calculated from the overall $(n + 1)$ observations, one observes that CLE in (4.4) and (4.5) are very close to the estimators in (4.1) and (4.2).

It can be easily verified that the functions g, $\partial g/\partial \alpha$, $\partial g/\partial \lambda$ and $\partial^2 g/\partial \alpha \, \partial \lambda$ satisfy all the regularity conditions of theorem (3.1) in Klimko and Nelson (1978). Consequently the CLS in (4.4) and (4.5) are strongly consistent. Moreover, if $E(|\varepsilon_t|^3) < \infty$, which is the case for a Poisson distribution, then by theorem (3.2) of Klimko and Nelson, $(\hat{\alpha}, \hat{\lambda})$ are asymptotically normally distributed as

$$n^{1/2}(\hat{\theta} - \theta^0) \sim \mathrm{MVN}(0, \, V^{-1}WV^{-1})$$

where $\hat{\theta} = (\hat{\alpha}, \hat{\lambda})'$ and $\theta^0 = (\alpha^0, \lambda^0)'$ denotes the 'true' value of the parameters. V is a 2×2 matrix with elements

$$V_{ij} = E\left(\frac{\partial g(\theta^0, X_{t-1})}{\partial \theta_i} \cdot \frac{\partial g(\theta^0, X_{t-1})}{\partial \theta_j} \right) \qquad i, j = 1, 2$$

and the elements of W have the form

$$W_{ij} = E\left(u_t^2(\theta^0) \frac{\partial g(\theta^0, X_{t-1})}{\partial \theta_i} \cdot \frac{\partial g(\theta^0, X_{t-1})}{\partial \theta_j} \right) \qquad i, j = 1, 2$$

with

$$u_t(\theta^0) = X_t - g(\theta^0, X_{t-1}).$$

4.3. *The maximum likelihood estimators*

The likelihood function of a sample of $(n + 1)$ observations from the INAR(1) process can be written as

$$L(x, \alpha, \lambda) = \left(\prod_{t=1}^{n} P_t(x_t) \right) \frac{[\lambda/(1 - \alpha)]^{x_0}}{x_0!} \exp[-\lambda/(1 - \alpha)] \qquad (4.6)$$

where

$$P_t(x) = \exp(-\lambda) \sum_{i=0}^{\min(x_{t-1}, x_t)} \frac{\lambda^{x-i}}{(x - i)!} \binom{x_{t-1}}{i} \alpha^i (1 - \alpha)^{x_{t-1}-i} \qquad \text{for } t = 1, 2, \ldots, n$$

and $x = (x_0, x_1, \ldots, x_n)$.

We shall consider first the conditional maximum likelihood estimators (CMLE) of α and λ when x_0 is given, in this case the conditional likelihood function is of the form:

$$L(x, \alpha, \lambda x_0) = \prod_{t=1}^{n} P_t(x_t).$$

Sprott (1983) described a method for maximizing $L(x, \alpha, \lambda/x_0)$ with respect to α and λ when $P_t(x)$ is independent of t, i.e. the index parameter of the observations. It is not difficult to see that the same procedure can be used in the present case. Following Sprott's approach one finds that the derivatives of the log conditional likelihood function with respect to λ and α are

$$S_\lambda = \sum_{t=1}^{n} \frac{\partial \log P_t(x_t)}{\partial \lambda} = \sum_{t=1}^{n} P_t(x_t - 1)/P_t(x_t) - n = \sum_{t=1}^{n} H(t) - n \qquad (4.7)$$

and

$$S_\alpha = \sum_{t=1}^{n} \frac{\partial \log P_t(x_t)}{\partial \alpha} = \sum_{t=1}^{n} [(x_t - \alpha x_{t-1}) - \lambda H(t)]/\alpha(1 - \alpha). \qquad (4.8)$$

To obtain the CMLE we set $S_\lambda = 0$ and $S_\alpha = 0$, which from (4.8) yields

$$\sum x_t = \alpha \sum x_{t-1} + n\lambda. \qquad (4.9)$$

Equation (4.9) can be used to eliminate one of the parameters, say α, then S_λ in (4.7) can be written as a function of λ only and hence CMLE can be found by iterating on S_λ. It should be noted that if one uses (4.9) to eliminate λ, then one would obtain an estimator of λ in terms of α which is of the same form as in (4.5). However, nothing can be said analytically in comparing the estimator α in (4.4) with that which would result from minimizing (4.8).

Now considering the unconditional likelihood function as given by (4.6), it can be seen that the derivatives of the log likelihood function with respect to λ and α are

$$S_\lambda^0 = \frac{\partial \log L}{\partial \lambda} = S_\lambda + \frac{x_0}{\lambda} - \frac{1}{1 - \alpha} \qquad (4.10)$$

and

$$S_\alpha^0 = \frac{\partial \log L}{\partial \alpha} = S_\alpha + \frac{1}{1 - \alpha}\left(x_0 - \frac{\lambda}{1 - \alpha}\right). \qquad (4.11)$$

The MLE can be found by setting $S_\lambda^0 = 0$ and $S_\alpha^0 = 0$, from which we obtain

$$\lambda = \left(\sum_{t=0}^{n} x_t - \alpha \sum_{t=1}^{n} x_t\right)\bigg/\left(n + \frac{1 + \alpha}{1 - \alpha}\right). \qquad (4.12)$$

Equation (4.12) can be used to eliminate λ in (4.11), then the MLE of α can be found by iterating on this equation. It is clear from equations (4.10), (4.11) and (4.12) that the MLE are not easy to manipulate as in the case of the CMLE.

The estimated asymptotic covariance matrix for the estimates of the parameters can be found by calculating the second derivative of the likelihood function. However, as Sprott remarked, estimates of the parameters in a convolution in general are highly correlated and consequently inferences about the parameters are highly interdependent.

In deriving the MLE we have assumed that (ε_t) has a Poisson distribution. It should be mentioned that Sprott (1983) remarked that the procedure can be applied for finding the MLE for the parameters of a convolution of two members of the power-series distribution family. Consequently this method can be modified in a similar manner to obtain the MLE for the parameters of the INAR(1) process for the case in which (ε_t) does not necessarily have a Poisson distribution.

To get some feeling for the relative merits of each of the three methods of estimation, we have done a simulation for the case in which an element of x_{t-1} has probability of survival α and the error term (ε_t) has a Poisson distribution with parameter λ. The initial value of the process (x_0) was taken to be the expected value of the process $[\lambda/(1 - \alpha)]$. To carry out the simulation the IMSL library is used to generate the number of successes from the binomial distribution with parameters (x_{t-1}, α) and from the Poisson with parameter (λ), with DSEED $= 36455$ (completely random). The simulation experiment was performed for samples of size $n = 50, 75, 100$ and 200 and for different values of the parameters $\alpha = 0.1(0.2)0.9$ and $\lambda = 1.0(0.5)3.0$. Estimates of the parameters α and λ in each sample were computed by utilizing equation (4.9) to eliminate the parameter λ. The CMLE of α was then found by iterating on S_α in equation (4.8). The subroutine ZBRENT in the IMSL library was used to find the zero solution of S_α. The computations were done on the IBM 3033 at King Saud University Computer Center. 200 replications were made on each sample, for every possible combination of the parameters α and λ, and the average bias and mean square error of the parameters' estimates were calculated.

The simulation results for the case $\lambda = 1$, which are typical, are presented in Tables I and II which show, respectively, the bias and the mean square error for different values of α and the various sample sizes which have been considered.

It can be seen from Table I that:

(i) Out of the 20 cases which are the different combinations of the values of α and the sample size, $\hat{\alpha}$ is biased down and $\hat{\lambda}$ is biased up in 17 cases of the CML,

TABLE I

Bias Results of Different Estimation Methods for the INAR(1) Model ($\lambda = 1.0$)
based on 200 Replications

n	α	Bias (α)			Bias (λ)		
		Y–W	CLS	CML	Y–W	CLS	CML
50	0.1	−0.0009	0.0002	0.0039	0.0044	0.0029	−0.0023
	0.3	−0.0527	−0.0481	−0.0316	0.0627	0.0562	0.0306
	0.5	−0.0649	−0.0556	−0.0162	0.1049	0.0874	0.0082
	0.7	−0.0766	−0.0647	−0.0177	0.2285	0.1892	0.0287
	0.9	−0.1054	−0.0786	−0.0048	1.0030	0.7424	0.0178
75	0.1	−0.0027	−0.0017	−0.0017	0.0015	0.0003	−0.0038
	0.3	−0.0382	−0.0346	−0.0254	0.0403	0.0351	0.0209
	0.5	−0.0377	−0.0322	−0.0075	0.0652	0.0544	0.0060
	0.7	−0.0555	−0.0464	−0.0104	0.1859	0.1552	0.0322
	0.9	−0.0703	−0.0548	−0.0013	0.6959	0.5441	−0.0003
100	0.1	−0.0075	−0.0069	0.0059	0.0026	0.0019	0.0002
	0.3	−0.0270	−0.0238	−0.0161	0.0275	0.0230	0.0123
	0.5	−0.0309	−0.0259	−0.0051	0.0652	0.0556	0.0138
	0.7	−0.0361	−0.0292	−0.0035	0.1158	0.0931	0.0018
	0.9	−0.0491	−0.0391	−0.0011	0.4858	0.3853	0.0091
200	0.1	−0.0062	−0.0056	−0.0035	0.0023	0.0017	−0.0007
	0.3	−0.0106	−0.0092	−0.0054	0.0084	0.0065	0.0010
	0.5	−0.0128	−0.0102	−0.0026	0.0219	0.0168	0.0018
	0.7	−0.0222	−0.0184	−0.0025	0.0728	0.0600	0.0062
	0.9	−0.0312	−0.0262	−0.0017	0.2989	0.2490	0.0036

19 cases of the CLS and 20 cases of the Yule–Walker estimates. This inverse relationship is to be expected from the equations defining the estimators (equations (4.2), (4.5) and (4.9)). However, the possibility that the two estimators $\hat{\alpha}$ and $\hat{\lambda}$ have the same bias sign can occur only if the value of x_n exceeds the mean of the process $\lambda/(1 - \alpha)$ in many replications. This can be seen from the following relation

$$\hat{\lambda} - \lambda = [\lambda/(1 - \alpha) - x_n]/n - (\hat{\alpha} - \alpha).$$

(ii) The magnitude of the biases of $\hat{\alpha}$ and $\hat{\lambda}$ increase with the increase in α in both the Yule–Walker and CLS estimates. The increase in the bias of $\hat{\lambda}$ is much higher (about eight times) than that of $\hat{\alpha}$. In contrast, the biases of the CML do not show such a tendency for increase in the bias. This observation is illustrated graphically in Figures 3(a) and 3(b) which show, respectively, the biases in $\hat{\alpha}$ and $\hat{\lambda}$ for the three methods of estimation based on a sample of size 75 and values of $\alpha = 0.1(0.1)0.9$.

(iii) In both the Yule–Walker and the CLS estimates, and to a lesser extent, in the CML estimates, the amount of bias is reciprocally related to the sample size. It can be seen that, by doubling the sample size, the amount of bias of each of $\hat{\alpha}$ and $\hat{\lambda}$ would be reduced by one-half in both the Yule–Walker and CLS estimates.

TABLE II

Mean Square Error Results of Different Estimation Methods for the INAR(1) Model ($\lambda = 1.0$) based on 200 Replications

n	α	MSE (α)			MSE (λ)		
		Y–W	CLS	CML	Y–W	CLS	CML
50	0.1	0.0116	0.0120	0.0142	0.0406	0.0405	0.0417
	0.3	0.0214	0.0216	0.0213	0.0663	0.0663	0.0605
	0.5	0.0223	0.0220	0.0137	0.1075	0.1061	0.0689
	0.7	0.0187	0.0180	0.0062	0.2217	0.2108	0.0656
	0.9	0.0197	0.0143	0.0007	1.8963	1.3670	0.0668
75	0.1	0.0074	0.0077	0.0083	0.0261	0.0262	0.0265
	0.3	0.0143	0.0145	0.0132	0.0397	0.0399	0.0349
	0.5	0.0155	0.0154	0.0098	0.0725	0.0718	0.0508
	0.7	0.0118	0.0113	0.0040	0.1476	0.1397	0.0473
	0.9	0.0100	0.0081	0.0004	1.0564	0.8718	0.0331
100	0.1	0.0063	0.0064	0.0067	0.0177	0.0177	0.0171
	0.3	0.0099	0.0100	0.0087	0.0301	0.0301	0.0274
	0.5	0.0095	0.0094	0.0055	0.0484	0.0479	0.0316
	0.7	0.0068	0.0064	0.0026	0.0875	0.0824	0.0251
	0.9	0.0055	0.0048	0.0003	0.5384	0.4593	0.0211
200	0.1	0.0039	0.0040	0.0041	0.0094	0.0094	0.0095
	0.3	0.0058	0.0059	0.0046	0.0158	0.0159	0.0136
	0.5	0.0048	0.0048	0.0022	0.0233	0.0232	0.0133
	0.7	0.0033	0.0031	0.0012	0.0389	0.0368	0.0133
	0.9	0.0023	0.0020	0.0001	0.2279	0.2000	0.0115

Regarding the MSE results in Table II it can be seen that:

(i) Whereas the MSE of $\hat{\lambda}$ in the Yule–Walker and CLS estimates increases with the increase in α, this is not the case for the MSE of $\hat{\alpha}$. In contrast the MSE of $\hat{\alpha}$ and $\hat{\lambda}$ in the CML estimates seems to be unaffected by the increase in the value of α. This remark can be seen from Figures 4(a) and 4(b) which show, respectively, the MSE for $\hat{\alpha}$ and $\hat{\lambda}$ for the three methods of estimation based on a sample of size 75 with $\alpha = 0.1(0.1)0.9$.

(ii) In the three methods of estimation, the MSE of each parameter is recipro-cally related to the sample size. The pace of decrease in the MSE as a result of increase in the sample size is similar for all methods (the MSE would be reduced by about 50% if the sample size is doubled).

In comparing the three methods of estimation in the light of the results of the simulation experiment, it seems that the CML method is worth the effort. It can be seen that the biases of $\hat{\alpha}$ and $\hat{\lambda}$ in the CML estimates are about 60% and 52%, respectively, of the corresponding Yule–Walker biases when $\alpha = 0.3$. These per-centages decrease monotonically with the increase in the value of α, reaching about 5% and 1% when $\alpha = 0.9$. Also, the gain in terms of the bias of the CML estimates over the Yule–Walker estimates increases with the increase in the sample size. The same remarks hold in comparing the biases of $\hat{\alpha}$ and $\hat{\lambda}$ in the
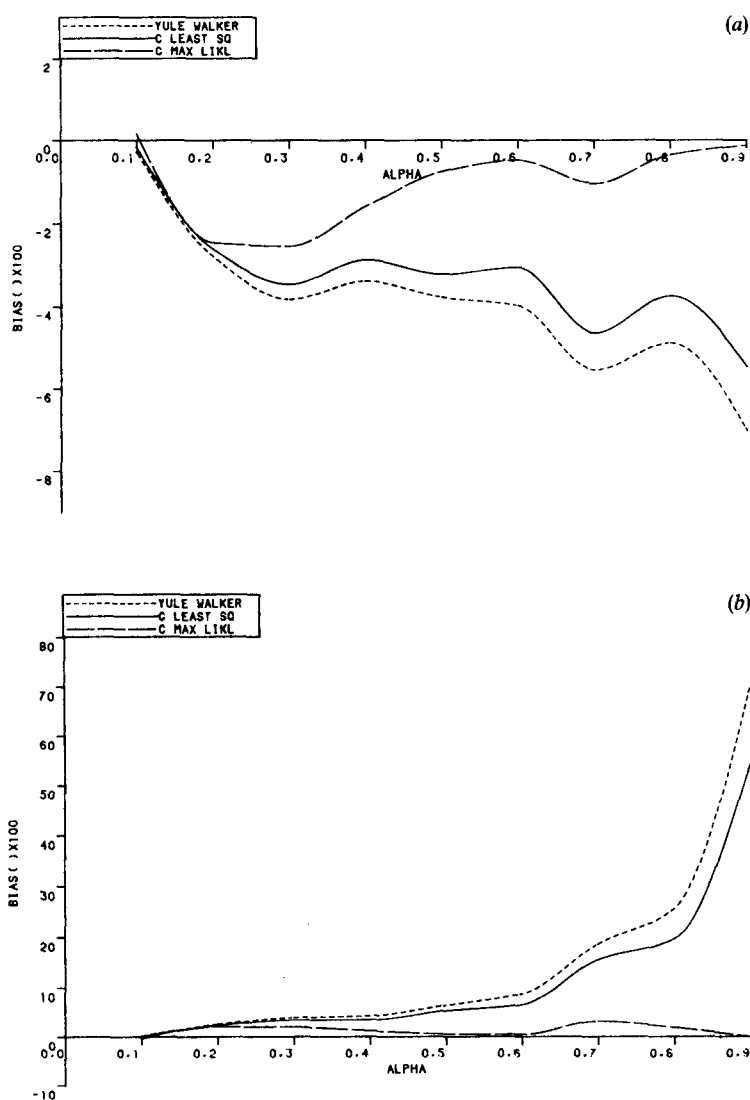
FIGURE 3. (a) Bias (α) results based on sample size 75 over 200 replications. (b) Bias (λ) results based on sample size 75 over 200 replications.

CML estimates with those of the CLS estimates. The percentages in this case are about 70% for the bias in α and about 50% for the bias of $\hat{\lambda}$ when $\alpha = 0.3$. In terms of the MSE, it can be seen that the MSE of $\hat{\alpha}$ and $\hat{\lambda}$ in the CML methods are around 90% of the corresponding MSE of the Yule–Walker and CLS estimates, when $\alpha = 0.3$. This percentage decreases to about 5% when $\alpha = 0.9$.

Regarding the case $\alpha = 0.1$, it can be seen from Tables I and II that the gain, in terms of the bias and MSE, of the Yule–Walker estimates over those of the CLS

and CML methods, diminishes with the increase in the sample size. Actually, for a sample of size 100 and over, estimates in the last two methods outperform those of the Yule–Walker in terms of the bias. The results of the simulation experiment presented here might cast some doubt upon the reliability of the CML estimates for samples of size 50, 75 and when $\alpha = 0.10$. To alleviate such a doubt it should be mentioned that, since the mean of the process is near 1 when $\alpha = 0.10$, the
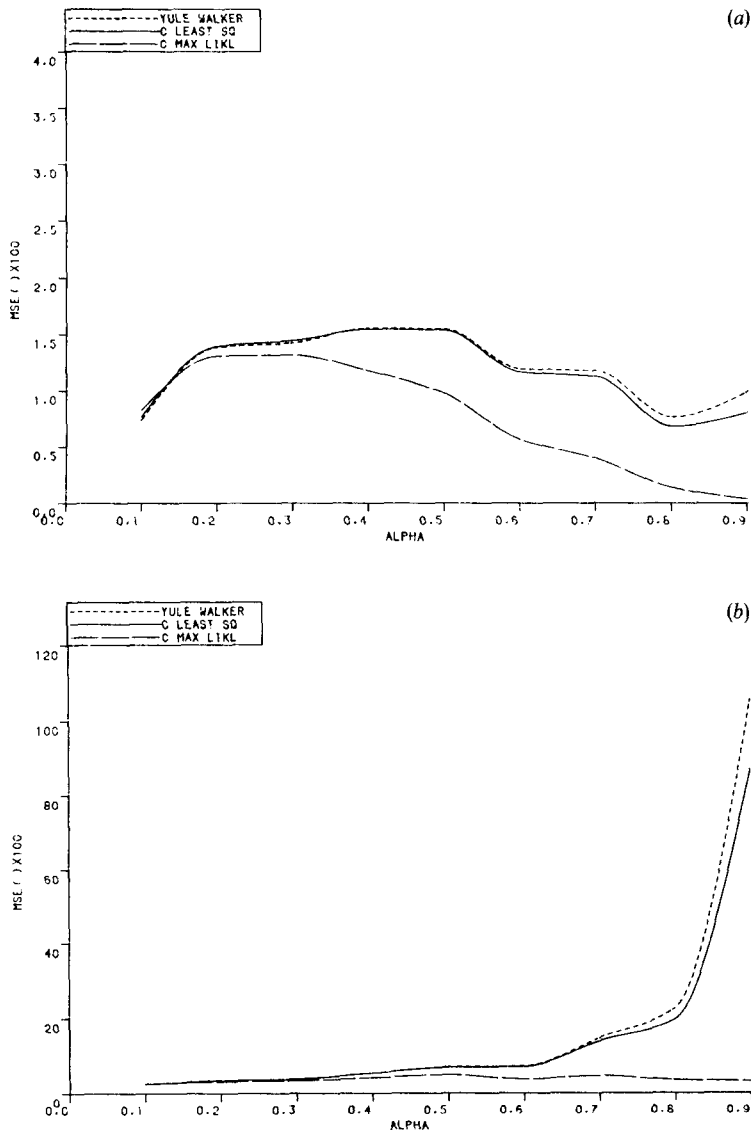

FIGURE 4. (a) MSE (α) results based on sample size 75 over 200 replications. (b) MSE (λ) results based on sample size 75 over 200 replications.

generated sample path might contain many zero values. Consequently, the zero solution of S_α (in equation (4.8)) within the admissible range of α is taken, by the subroutine used above, to be zero. Also, it should be pointed out that the results of a Monte Carlo study done on the standard AR(1) process showed that the Yule–Walker estimates might have a smaller bias and/or MSE than the least squares or the maximum likelihood when α is near zero (see e.g. Dent and Min, 1978).

In summary, the conclusion which can be drawn from the results of the simulation experiment is that if a ranking were to be made among the three methods of estimation, the CML would be first, followed by the CLS, and then the Yule–Walker method. The magnitude of gain, in terms of the bias and MSE, in the CML estimates compared with the other two methods makes it worth the extra calculations.

6. SUMMARY AND CONCLUSIONS

It is intuitive, in modelling counting process with dependence, to consider that a realization of the process (X_t) is composed of survivals of the elements of the process in the previous period(s) (x_{t-1}) in addition to the new arrival process (ε_t). The simplicity of the proposed model in this paper makes it applicable to many real life situations in which a model for counts is of interest. The correlation structure and the distributional properties of the model are discussed and compared with those of the standard AR(1) process. Three methods for estimating the parameters of the model are considered under the assumption that the error sequences (ε_t) have the Poisson distribution, and the relative merits of these methods are compared via a simulation study. However, the same estimation procedures are still applicable to other distributions for the errors.

Extension of the INAR(1) to higher lag dependence is straightforward. Further, a possible generalization of the model is to consider the case in which the probability of survival for new comers (ε_t) might differ from that of their predecessors in the system during an adjustment period, which might be due to drop-out during a training period. Generalization of the model to cover these and similar aspects is the subject of a forthcoming paper.

REFERENCES

ALI KHAN, M. S. (1986) An infinite dam with random withdrawal policy. *Adv. Appl. Prob.* 18, 933–51.
BONDENSSON, L. (1981) Discussion of a paper by D. R. Cox: Statistical analysis of time series: some recent developments. *Scand. J. Statist.* 8, 93–115.

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

Dent, W. and Min, A. (1978) A Monte Carlo study of autoregressive integrated moving average process. *J. Econometrics* 7, 23–55.

Gaver, D. P. and Lewis, P. A. W. (1980) First-order autoregressive gamma sequences and point processes. *Adv. Appl. Prob.* 12, 727–45.

Jacobs, P. A. and Lewis, P. A. W. (1978a) Discrete time series generated by mixtures I: correlation and runs properties. *J. Roy. Statist. Soc.* B 40, 94–105.

—— (1978b) Discrete time series generated by mixtures II: asymptotic properties. *J. Roy. Statist. Soc.* B 40, 222–8.

—— (1983) Stationary discrete autoregressive moving average time series generated by mixtures. *J. Time Series Anal.* 4, 19–36.

Klimko, L. A. and Nelson, P. I. (1978) On conditional least squares estimation for stochastic processes. *Ann. Statist.* 6, 629–42.

Sprott, D. A. (1983) Estimating the parameters of a convolution by maximum likelihood. *J. Amer. Statist. Ass.* 78, 457–60.

Stuetal, F. W. and Van Harn, K. (1979) Discrete analogues of self-decomposability and stability. *Ann. Prob.* 7, 893–9.

Van Harn, K. (1978) *Classifying Infinitely Divisible Distributions by Functional Equations.* Amsterdam: Mathematisch Centrum.