TA gl.

FORSCHUNG - AUSBILDUNG - WEITERBILDUNG

Bericht Nr. 95

MULTIVARIATE FIRST-ORDER INTEGER-VALUED AUTOREGRESSIONS

Jürgen Franke and T. Subba Rao

UNIVERSITÄT KAISERSLAUTERN
Fachbereich Mathematik
Postfach 3049

D-67653 Kaiserslautern

September 1993

Jürgen Franke and T. Subba Rao

Multivariate first-order integer-valued autoregressions

Address of corresponding author: Universität Kaiserslautern
Fachbereich Mathematik
Postfach 3049
67653 Kaiserslautern
Germany

1 Introduction

Linear models like autoregressions or ARMA-processes have primarily been designed for real-valued and, in particular, Gaussian time series. They do not seem to be adequate tools for integer-valued processes, e.g. daily or annual counts, if the observations are contained in a small finite set with high probability. A flexible parametric class of parametric models for time series with values in $\{0, 1, 2, \ldots\}$ are the integer-valued autoregressive (INAR) processes which have been discussed by Al-Osh and Alzaid (1987) and Du and Li (1991). The same type of model has been introduced under a different name by McKenzie (1985-1988) who derived various properties for the case of specific marginal distributions and extended the concept to ARMA-like processes. Let us also remark that INAR-processes of order 1 are special Galton-Watson processes with immigration as studied by Seneta (1969), Venkataraman (1982) and Venkataraman and Nanthi (1982).

Franke and Seligmann (1993) describe the successful application of a particular kind of INAR-process of order 1 to modelling daily counts of epileptic seizure. As a theoretical basis for the model and the estimation procedure they gave conditions for the strict stationarity of INAR(1)-processes and investigated the asymptotic properties of conditional maximum likelihood (CML) estimates for the model parameters.

In this paper we generalize the concept of integer-valued autoregressions from the univariate to the multivariate case and discuss stationarity conditions and the properties of CML-estimates for models of order 1. Such models and estimation procedures are useful for fitting time series of vectors of counts. Some applications, which we have in mind, are to daily counts of epileptic seizures, distinguished with respect to type of seizure, and to annual counts of a deer population distinguished with respect to sex and age class. Multivariate INAR(1)-models have one great advantage: the integer-valued process $X(t) = A \circ X(t-1) + \varepsilon(t)$. That fact follows easily from the rules for calculating first and second moments given in Lemma 1 below. Therefore, INAR-processes retain some of the properties of the familiar autoregressions while allowing for the discreteness of the data.

Apart from applications to genuine multivariate data our discussion of INAR(1)-processes in higher dimensions are motivated by the fact that scalar INAR(M)-processes, as discussed by Du and Li (1991), can be written as M-variate INAR(1)-processes. In the latter form, they have the Markov property, and we are able to apply directly the asymptotic theory of Billingsley (1961) to CML-estimates of INAR(M)-parameters.

2 Multivariate integer-valued autoregressions of order 1

To characterize the class of time series models, which we are interested in, we have to introduce some notation:

$$p \circ U = \sum_{j=1}^{U} Y_j$$

where U is a random variable with values in $\mathbb{N}_0 = \{0,1,2,\ldots\}$, $0 \le p \le 1$, and $Y_1, Y_2, \ldots$ are i.i.d. Bernoulli variables, independent of U, with

$$p = pr(Y_j = 1) = 1 - pr(Y_j = 0).$$

We call $Y_1, Y_2, \ldots$ the counting series of $p \circ U$, and we remark that, given U, $p \circ U$ has a binomial distribution with parameters (U, p). In general, for $M \ge 1$, let A be a $M \times M$-matrix with entries a_{ij} satisfying $0 \le a_{ij} \le 1$ for $i,j = 1,\ldots,M$. Then, for a random vector X with values in $\mathbb{N}_0^M$ we define $A \circ X$ as the $\mathbb{N}_0^M$-valued random vector with i-th component

$$(A \circ X)_i = \sum_{j=1}^{M} a_{ij} \circ X_j \ , \quad i = 1,\ldots,M,$$

where we assume independence of the counting series of all $a_{ij} \circ X_j$, $i,j = 1,\ldots,M$. We stress that we use this definition for constant $X \equiv \underline{m} \in \mathbb{N}_0^M$, too.

A straightforward calculation shows $p \circ (q \circ U) \stackrel{d}{=} (pq) \circ U$, where $\stackrel{d}{=}$ stands for equal distributions, in dimension 1, and, then in dimension M:

$$A \circ (B \circ X) \stackrel{d}{=} (AB) \circ X \tag{1}$$

We need the following rules for calculating moments of $A \circ X$:

Lemma 1: *(i)* $\mathsf{E}(A \circ X) = A\, \mathsf{E}X$

(ii) If all the counting series of $A \circ X$ and $B \circ Z$ are independent,

$$\mathsf{E}\{(A \circ X)(B \circ Z)^T\} = A\ \mathsf{E}(XZ^T)\, B^T$$

(iii)
$$\mathsf{E}\{(A \circ X)(A \circ X)^T\} = A\ \mathsf{E}(XX^T)\, A^T + diag\ (C\, \mathsf{E}X)$$

where C is the $M \times M$-matrix with entries $C_{ij} = a_{ij}(1 - a_{ij})$, $i,j = 1,\ldots,M$, and, for $Z \in \mathbb{R}^M$, diag (Z) *is the diagonal matrix with entries $Z_1,\ldots,Z_M$ in the diagonal.*

Proof: (i) and (ii) follow directly from the one-dimensional relations given by

$$\mathsf{E}(p \circ U) = p\, \mathsf{E}U \ , \quad \mathsf{E}\{(p \circ U)(q \circ V)\} = pq\, \mathsf{E}(UV),$$

assuming independent counting series of $p \circ U$ and $q \circ V$. We prove (iii) for each component separately, i.e. for $k, l = 1,\ldots,M$, we consider

$$[\mathsf{E}\{(A \circ X)(A \circ X)^T\}]_{k,l} = \sum_{i,j=1}^{M} \mathsf{E}\{(a_{ki} \circ X_i)(a_{lj} \circ X_j)\}$$

If $k \neq l$ or $i \neq j$, $a_{ki} \circ X_i$ and $a_{lj} \circ X_j$ are conditionally independent binomial random variables given X_i and X_j, and

$$\begin{aligned} \mathsf{E}\{(a_{ki} \circ X_i)(a_{lj} \circ X_j)\} &= \mathsf{E}[\mathsf{E}\{(a_{ki} \circ X_i)(a_{lj} \circ X_j)|X_i, X_j\}] \\ &= \mathsf{E}[a_{ki}\, X_i\, a_{lj}\, X_j] \\ &= a_{ki}\, a_{lj}\, \mathsf{E}(X_i X_j)\,. \end{aligned}$$

But if $k = l$ and $i = j$, we have

$$\begin{aligned} \mathsf{E}(a_{ki} \circ X_i)^2 &= \mathsf{E}[\mathsf{E}\{(a_{ki} \circ X_i)^2|X_i\}] \\ &= \mathsf{E}[X_i\, a_{ki}(1 - a_{ki}) + X_i^2\, a_{ki}^2] \\ &= a_{ki}^2\, \mathsf{E}\, X_i^2 + a_{ki}(1 - a_{ki})\, \mathsf{E}\, X_i \end{aligned}$$

Therefore, with $\delta_{kl} = 1$ if $k = l$ and $= 0$ else,

$$[\mathsf{E}\{(A \circ X)(A \circ X)^T\}]_{k,l} = \sum_{i,j=1}^{M} a_{ki}\, a_{lj}\, \mathsf{E}(X_i X_j) + \delta_{kl} \sum_{i=1}^{M} a_{ki}(1 - a_{ki})\, \mathsf{E}\, X_i$$

■

Definition 2.1: *A $\mathbb{N}_0^M$-valued time series $\{X(t),\ -\infty < t < \infty\}$ is called a M-variate INAR(1)-process if*

(2) $$X(t) = A \circ X(t-1) + \varepsilon(t)\,, \quad -\infty < t < \infty\,,$$

for some $M \times M$-matrix A with entries $0 \leq a_{ij} \leq 1$, $i, j = 1, \ldots, M$, and i.i.d. $\mathbb{N}_0^M$-valued random variables $\varepsilon(t)$, $-\infty < t < \infty$.

Let $$q^{\beta}(\underline{\mathbf{k}}) = pr(\varepsilon(t) = \underline{\mathbf{k}})\,, \quad \underline{\mathbf{k}} \geq \underline{\mathbf{0}}\,,$$

denote the weights of the law of the innovations $\varepsilon(t)$, which we assume to depend on some parameter $\beta \in \mathcal{B}$, where $\mathcal{B}$ is an open subset of $\mathbb{R}^d$. Here and in the following, we write $\underline{\mathbf{m}} \geq \underline{\mathbf{n}}$ for $\underline{\mathbf{m}}, \underline{\mathbf{n}} \in \mathbb{Z}^M$ if $m_j \geq n_j$, $j = 1, \ldots, M$.

$\{X(t)\}$ is a Markov chain with states in $\mathbb{N}_0^M$ and transition probabilities

$$\begin{aligned} P_{\vartheta}(\underline{\mathbf{m}}, \underline{\mathbf{n}}) &= pr(X(t) = \underline{\mathbf{n}}|X(t-1) = \underline{\mathbf{m}}) \\ &= \sum_{\underline{\mathbf{k}} \leq \underline{\mathbf{n}}} pr(A \circ \underline{\mathbf{m}} = \underline{\mathbf{n}} - \underline{\mathbf{k}})\, q^{\beta}(\underline{\mathbf{k}})\,, \quad \underline{\mathbf{m}}, \underline{\mathbf{n}} \geq 0\,. \end{aligned}$$

Here, ϑ stands for the unknown parameters of the process $\{X(t)\}$, i.e. in general for β and for some $\alpha \in \mathbb{R}^s$, $s \leq M^2$, determining the entries a_{ij} of A. It is admissible that all the a_{ij}, $i,j = 1,\ldots,M$, appear as separate components of α, i.e. as free parameters in their own right, but they also may be given functions of some lower dimensional quantity. For later use, we extend the definition of $P_\vartheta(\underline{\mathrm{m}},\underline{\mathrm{n}})$ to $\mathbb{Z}^M \times \mathbb{Z}^M$ by setting $P_\vartheta(\underline{\mathrm{m}},\underline{\mathrm{n}}) = 0$ if at least one component of $\underline{\mathrm{m}}$ or of $\underline{\mathrm{n}}$ is negative.

The existence of a stationary solution of (2) depends on the largest eigenvalue of the non-negative matrix A. By a variant of the Perron-Frobenius-Theorem (Basilevsky, 19.., ...), there exists an eigenvalue η_1 of A such that $\eta_1 \geq |\eta_j|$ for all other eigenvalues η_j of A, and the eigenvector $\underline{\mathbf{z}}$ corresponding to η_1 satisfies $\underline{\mathbf{z}} \geq \underline{\mathbf{0}}$. If A is even positive, i.e. $a_{ij} > 0$ for $i,j = 1,\ldots,M$, we have strict inequalities: $\eta_1 > |\eta_j|$, $j \neq 1$, and $\underline{\mathbf{z}} > \underline{\mathbf{0}}$.

<u>Theorem 1:</u> *Let $\{X(t)\}$, satisfying (2), be an irreducible, aperiodic Markov chain on $\mathbb{N}_0^M$. If $\mathsf{E}||\varepsilon(t)|| < \infty$ and if the largest eigenvalue η_1 of A is less than 1, then there exists a strictly stationary multivariate INAR(1)-process satisfying (2).*

Proof:

1. Let $P_\vartheta^{(t)}(\underline{\mathrm{m}},\underline{\mathrm{n}}) = pr(X(t) = \underline{\mathrm{n}}|X(0) = \underline{\mathrm{m}})$ denote the t-step transition probabilities. Using (1) repeatedly, we have

$$X(t) = A \circ (A \circ X(t-2)) + A \circ \varepsilon(t-1) + \varepsilon(t) = \ldots \stackrel{\mathrm{d}}{=} \sum_{j=0}^{t-1} A^j \circ \varepsilon(t-j) + A^t \circ X(0)$$

Using independence of the $\varepsilon(s)$ and nonnegativity of their coordinates we get

$$\begin{aligned} P_\vartheta^{(t)}(\underline{\mathbf{0}},\underline{\mathbf{0}}) &= pr(X(t) = \underline{\mathbf{0}}|X(0) = \underline{\mathbf{0}}) \\ &= pr\left(\sum_{j=0}^{t-1} A^j \circ \varepsilon(t-j) = \underline{\mathbf{0}}\right) \\ &= \prod_{j=0}^{t-1} pr\left(A^j \circ \varepsilon(t-j) = \underline{\mathbf{0}}\right) \end{aligned}$$

Let $\mu = \mathsf{E}\,\varepsilon(t)$. By Markov's inequality

$$\begin{aligned} pr(A^j \circ \varepsilon(s) \neq \underline{\mathbf{0}}) &= pr\left(\sum_{i=1}^{M} \left(A^j \circ \varepsilon(s)\right)_i \geq 1\right) \\ &\leq \mathsf{E}\sum_{i=1}^{M} \left(A^j \circ \varepsilon(s)\right)_i \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{M} \left(\mathsf{E}\, A^j \circ \varepsilon(s)\right)_i \\
&= \sum_{i=1}^{M} \left(A^j \mu\right)_i = |A^j \mu|
\end{aligned}
$$

applying Lemma 1 and using the notation $|\underline{\mathbf{z}}| = \sum_{i=1}^{M} |z_i|$ for $\underline{\mathbf{z}} \in \mathbb{R}^{M}$. Now, let $\eta_1, \ldots, \eta_M$ be the eigenvalues of A, let $\underline{\mathbf{z}}^1, \ldots, \underline{\mathbf{z}}^M$ be a corresponding set of linearly independent eigenvectors, and let $\mu = \alpha_1 \underline{\mathbf{z}}^1 + \ldots + \alpha_M \underline{\mathbf{z}}^M$ be the representation of μ in terms of that basis. By Theorem 5.6 of Basilevsky (1983),

$$
A^j\, \underline{\mathbf{z}}^k = \eta_k^j \underline{\mathbf{z}}^k\,,\ j \geq 0,\ k = 1, \ldots, M,
$$

and, as $\eta_1 \geq |\eta_i|$ for all $i \geq 2$,

$$
\begin{aligned}
|A^j \mu| &= \left|\sum_{k=1}^{M} \alpha_k\, \eta_k^j\, \underline{\mathbf{z}}^k\right| \leq \sum_{k=1}^{M} |\eta_k|^j |\alpha_k\, \underline{\mathbf{z}}^k| \\
&\leq \eta_1^j\, c(\mu)
\end{aligned}
$$

with $c(\mu) = |\alpha_1\, \underline{\mathbf{z}}^1| + \ldots + |\alpha_M\, \underline{\mathbf{z}}^M|$. Therefore,

$$
pr\left(A^j \circ \varepsilon(s) = \underline{\mathbf{0}}\right) \geq 1 - |A^j \mu| \geq 1 - c(\mu)\eta_1^j
$$

Now, we choose $\tau > 0$ such that $c(\mu)\eta_1^\tau < 1$, and we have

$$
P_\vartheta^{(t)}(\underline{\mathbf{0}}, \underline{\mathbf{0}}) \geq \prod_{j=0}^{\tau-1} pr\left(A^j \circ \varepsilon(t-j) = \underline{\mathbf{0}}\right) \cdot \prod_{j=\tau}^{t-1} \left(1 - c(\mu)\eta_1^j\right)
$$

The right-hand side is positive and decreasing for $t \to \infty$. As, using the Taylor expansion of $\log(1 - x)$,

$$
\begin{aligned}
\log \prod_{j=\tau}^{t-1} \left(1 - c(\mu)\eta_1^j\right) &= -\sum_{k=1}^{\infty} \frac{1}{k} c(\mu)^k \sum_{j=\tau}^{t-1} \eta_1^{jk} = -\sum_{k=1}^{\infty} \frac{1}{k} c(\mu)^k \frac{\eta_1^{k\tau} - \eta_1^{kt}}{1 - \eta_1^k} \\
&\geq -\sum_{k=1}^{\infty} \frac{1}{k} c(\mu)^k \frac{\eta_1^{k\tau}}{1 - \eta_1} = \frac{1}{1 - \eta_1} \log(1 - c(\mu)\eta_1^\tau) > -\infty
\end{aligned}
$$

for all $t > \tau$, $P^{(t)}(\underline{\mathbf{0}}, \underline{\mathbf{0}})$ is also bounded from below by a positive number for $t \to \infty$. Therefore, we finally conclude

$$
\lim_{t \to \infty} P^{(t)}(\underline{\mathbf{0}}, \underline{\mathbf{0}}) > 0.
$$

2. From 1. we have immediately

$$\sum_{t=0}^{\infty} P^{(t)}(\underline{0}, \underline{0}) = \infty ,$$

i.e. $\underline{0}$ is a recurrent state. Let τ_0 be its mean recurrence time. By Theorem 1.2.2. of Rosenblatt (1971),

$$\frac{1}{\tau_0} = \lim_{t \to \infty} P^{(t)}(\underline{0}, \underline{0}) ,$$

and, by 2. $\tau_0 < \infty$ and, therefore, $\underline{0}$ is a positive recurrent state. This implies the existence of a strictly stationary solution of (2) by Theorem 1.2.1 of Rosenblatt (1971) and the remarks before the statement of this result.

■

Except for degenerate situations, the assumption of irreducibility and aperiodicity are satisfied. For instance, we have

Lemma 2: *Let* $0 < a_{ij} < 1$, $i, j = 1, \ldots, M$ *and* $0 < q^{\beta}(\underline{0}) < 1$. *Then, any solution of (2) is an irreducible and aperiodic Markov chain on* $\mathbb{N}_0^M$.

Proof:
Let $\underline{e}_1, \ldots, \underline{e}_M$ denote the unit vectors of $\mathbb{R}^M$. As $0 < a_{ij} < 1$, $i, j = 1, \ldots, M$, and $0 < q^{\beta}(\underline{0}) < 1$, we have

$$P_{\vartheta}(\underline{m}, \underline{0}) > 0 \text{ for all } \underline{m},\ P_{\vartheta}^{(2)}(\underline{0}, \underline{e}_j) > 0 \text{ for } j = 1, \ldots, M ,$$

$$P_{\vartheta}(\underline{m}, \underline{m} + \underline{e}_j) > 0 ,\ j = 1, \ldots, M, \text{ for all } \underline{m} \neq \underline{0},$$

which implies the irreducibility and aperiodicity of $\{X(t)\}$.

■

The assumption of Lemma 2 is not satisfied if $\{X(t)\}$ is the M-variate representation of a one-dimensional INAR(M)-process $U(t)$ as studied by Du and Li (1991):

$$U(t) = \sum_{k=1}^{M} \alpha_k \circ U(t-k) + \nu(t)$$

with $0 \leq \alpha_k \leq 1$, $k = 1, \ldots, M$, and i.i.d. $\mathbb{N}_0$-valued random variables $\nu(t)$. This process can be written as

(3) $$X(t) = A \circ X(t-1) + \varepsilon(t) \quad \text{with}$$

$$X(t) = (U(t), U(t-1), \ldots, U(t-M+1))^T ,\ \varepsilon(t) = (\nu(t), 0, \ldots, 0)^T , \text{ and}$$

$$A = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & & \alpha_M \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

__Lemma 3:__ *Let $0 < \alpha_k < 1$, $k = 1, \ldots, M$, and $0 < q^\beta(\underline{0}) = pr(\nu(t) = 0) < 1$. Then, any solution $\{X(t)\}$ of (3) is an irreducible and aperiodic Markov chain on $\mathbb{N}_0^M$.*

Proof: For $\underline{m} = (m_1, \ldots, m_M)^T \in \mathbb{N}_0^M$, let $\tilde{m} = (m_M, m_1, \ldots, m_{M-1})^T$. As $\alpha_1, \ldots, \alpha_{M-1} < 1$, $\alpha_M > 0$ and $q^\beta(\underline{0}) > 0$, we have $P_\vartheta(\underline{m}, \underline{\tilde{m}}) > 0$. Using this relation, $0 < \alpha_1, \ldots, \alpha_M < 1$ and $pr(\nu(t) = l) > 0$ for some $l \geq 1$, repeatedly, it is easy to see that

$$P_\vartheta^{(2M)}(\underline{m}, \underline{m} + \underline{e}_j) > 0 \ , \ j = 1, \ldots, M, \quad \text{for all} \quad \underline{m} \in \mathbb{N}_0^M \ ,$$

where $\underline{e}_1, \ldots, \underline{e}_M$ denote the unit vectors of $\mathbb{R}^M$. As also

$$P_\vartheta^{(M)}(\underline{m}, \underline{0}) > 0 \ \text{for all} \quad \underline{m} \in \mathbb{N}_0^M \ ,$$

the irreducibility of the Markov chain (3) follows. Then, $P_\vartheta(\underline{0}, \underline{0}) > 0$ implies immediately the aperiodicity of $\{X(t)]$ by the remarks of Rosenblatt (1974), ch. III c. ∎

3 Estimation of multivariate INAR(1)-parameters

In this chapter, we consider the problem of estimating the INAR(1)-parameters from a finite sample $X = (X(0), \ldots, X(N))$. We consider the conditional log-likelihood

$$\ell_N(X, \vartheta | X(0)) = \sum_{t=1}^{N} \log P_\vartheta \left(X(t-1), X(t) \right)$$

As discussed in Franke and Seligmann (1993), we prefer the conditional maximum likelihood (CML-)estimate $\hat{\vartheta} = (\hat{\alpha}, \hat{\beta})$ which maximizes $\ell_N(X, \vartheta | X(0))$ as a function of $\vartheta = (\alpha, \beta)$. Again, $\beta \in \mathcal{B} \subseteq \mathbb{R}^d$ determines the law of the innovations $\varepsilon(t)$, and we assume $\mathcal{B}$ to be open. $\alpha \in \mathbb{R}^s$ determines the coefficient matrix A of our INAR(1)-scheme. To keep the discussion as simple as possible we consider only the case where some of the entries of A are fixed and the others are completely unknown. In other words the components of α are just certain entries of A, say

(4a) $$\alpha_k = a_{i(k), j(k)} \ , \ k = 1, \ldots, s \ , \ \text{unknown} \quad ,$$

and the other entries of A are given constants:

(4b) $\quad a_{ij}$ unknown for $(i,j) \notin I = \{(i(k), j(k)),\ k = 1, \dots, s\}$.

We assume $0 < \alpha_k < 1$, $k = 1, \dots, s$ whereas the a_{ij}, $(i,j) \notin I$, may be 0 or 1. Therefore, the parameter ϑ which determines the transition probabilities satisfies

$$\vartheta = (\alpha, \beta) \in \Theta = (0,1)^s \times \mathcal{B} \subseteq \mathbb{R}^{s+d}$$

This framework includes the case where all a_{ij} are to be estimated from the data ($s = M^2$) as well as the case where A corresponds to a univariate INAR(M)-scheme as in (3) and its entries are 0 or 1 except for the first line ($s = M$).

We want to apply results of Billingsley (1961) on estimates of the parameters of Markov processes. For this purpose we have to impose some regularity conditions (C1) - (C6). We always assume that the stationarity condition of Theorem 1 is satisfied.

(C1) $\{\underline{\mathbf{k}}; q^\beta(\underline{\mathbf{k}}) > 0\}$ does not depend of β.

(C2) $\mathsf{E}||\varepsilon(t)||^3 = \sum_{\underline{\mathbf{k}} \geq \underline{\mathbf{0}}} ||\underline{\mathbf{k}}||^3 q^\beta(\underline{\mathbf{k}}) < \infty$.

(C3) For any $\underline{\mathbf{k}}$, $q^\beta(\underline{\mathbf{k}})$ is three times continuously differentiable with respect to β_u, $u = 1, \dots, d$, on $\mathcal{B}$.

(C4) For any $\beta' \in \mathcal{B}$ there exists a neighbourhood U of β' such that

$$\sum_{\underline{\mathbf{k}} \geq \underline{\mathbf{0}}} \sup_{\beta \in U} q^\beta(\underline{\mathbf{k}}) < \infty$$

$$\sum_{\underline{\mathbf{k}} \geq \underline{\mathbf{0}}} \sup_{\beta \in U} \left|\frac{\partial}{\partial \beta_u} q^\beta(\underline{\mathbf{k}})\right| < \infty\ , \quad u = 1, \dots, d,$$

$$\sum_{\underline{\mathbf{k}} \geq \underline{\mathbf{0}}} \sup_{\beta \in U} \left|\frac{\partial^2}{\partial \beta_u \partial \beta_v} q^\beta(\underline{\mathbf{k}})\right| < \infty\ , \quad u, v = 1, \dots, d,$$

(C5) For $u, v, \omega = 1, \dots, d$ and any $\beta' \in \mathcal{B}$ there exists a neighbourhood U of β' and functions $\Psi_u(\underline{\mathbf{n}})$, $\Psi_{uv}(\underline{\mathbf{n}})$, $\Psi_{uv\omega}(\underline{\mathbf{n}})$, $\underline{\mathbf{n}} \geq \underline{\mathbf{0}}$ (depending on β' and U), which increase with respect to the partial ordering of $\mathbb{N}_0^M$ induced by the definition of $\underline{\mathbf{k}} \leq \underline{\mathbf{n}}$, such that for all $\beta \in U$ and $\underline{\mathbf{k}} \leq \underline{\mathbf{n}}$ with nonvanishing $q^\beta(\underline{\mathbf{k}})$

$$\begin{aligned} \left|\frac{\partial}{\partial \beta_u} q^\beta(\underline{\mathbf{k}})\right| &\leq \Psi_u(\underline{\mathbf{n}})\, q^\beta(\underline{\mathbf{k}}) \\ \left|\frac{\partial^2}{\partial \beta_u \partial \beta_v} q^\beta(\underline{\mathbf{k}})\right| &\leq \Psi_{uv}(\underline{\mathbf{n}})\, q^\beta(\underline{\mathbf{k}}) \\ \left|\frac{\partial^3}{\partial \beta_u \partial \beta_v \partial \beta_\omega} q^\beta(\underline{\mathbf{k}})\right| &\leq \Psi_{uv\omega}(\underline{\mathbf{n}})\, q^\beta(\underline{\mathbf{k}}). \end{aligned}$$

Furthermore, we assume w.r.t. the stationary distribution of the multivariate INAR(1)-process $\{X(t)\}$

$$\mathsf{E}\, \Psi_u^3(X(1)) < \infty\,,\ \mathsf{E}\,\{||X(1)||\Psi_{uv}(X(2))\} < \infty\,,$$

$$\mathsf{E}\,\Psi_u(X(1))\,\Psi_{vw}(X(2)) < \infty\,,\ \mathsf{E}\,\Psi_{uvw}(X(1)) < \infty\,.$$

(C6) $\Sigma(\vartheta) = (\sigma_{uv}(\vartheta))_{u,v=1,\ldots,s+d}$, the Fisher information matrix given by

$$\sigma_{uv}(\vartheta) = \mathsf{E}\,(\frac{\partial}{\partial\vartheta_u}\log P_\vartheta(X(1),X(2))\,\frac{\partial}{\partial\vartheta_v}\log P_\vartheta(X(1),X(2))).$$

for $u,\ v = 1,\ldots,s+d$, is nonsingular.

Note that conditions (C1), (C4) and (C5) are automatically satisfied for any innovation law with bounded support, i.e. with only finitely many nonvanishing $q^\beta(\underline{\mathbf{k}})$.

Theorem 2: *Let $\{X(t)\}$ be a M-variate INAR(1)-process satisfying the stationarity condition of Theorem 1, and additionally (C1)-(C6). Then, the CML-estimate $\hat\vartheta = (\hat\alpha,\hat\beta)$ is asymptotically normal, i.e.*

$$\sqrt{N}(\hat\vartheta - \vartheta) \xrightarrow[\mathcal{L}]{} \mathcal{N}(\underline{0},\Sigma^{-1}(\vartheta)) \quad for \quad N\to\infty\,.$$

Furthermore, for $N\to\infty$,

$$2\{\ell_N(X,\hat\vartheta|X(0)) - \ell_N(X,\vartheta|X(0))\} \xrightarrow[\mathcal{L}]{} \chi^2_{s+d}$$

$$and \quad 2\{\ell_N(X,\hat\vartheta|X(0)) - \ell_N(X,\vartheta|X(0))\} - N(\hat\vartheta-\vartheta)^T\,\Sigma(\vartheta)(\hat\vartheta-\vartheta) \xrightarrow[\mathrm{P}]{} 0\,.$$

We postpone the proof of the theorem to the appendix. To illustrate the application of our result, we consider a simple example, a univariate INAR(2)-process:

(5a) $$U(t) = \alpha_1\circ U(t-1) + \alpha_2\circ U(t-2) + \nu(t)$$

with $0 < \alpha_1,\ \alpha_2 < 1$ and independent Poisson innovations $\nu(t)$ with common parameter $\beta > 0$. The bivariate INAR(1)-representation of this process is

(5b) $$X(t) = A\circ X(t-1) + \varepsilon(t)$$

$$\text{with}\ \ X(t) = \begin{pmatrix} U(t) \\ U(t-1) \end{pmatrix},\ \varepsilon(t) = \begin{pmatrix} \nu(t) \\ 0 \end{pmatrix},\ A = \begin{pmatrix} \alpha_1 & \alpha_2 \\ 1 & 0 \end{pmatrix}.$$

Calculating the eigenvalues of A, we see that the larger one is less than 1 iff $\alpha_1+\alpha_2<1$. By Theorem 1, the latter condition guarantees the existence of a stationary solution of (5a).

Franke and Seligmann (1993) pointed out that the Poisson innovations $\nu(t)$ satisfy the univariate version of conditions (C1)-(C5). As for the model (5b), $q^\beta(\underline{\mathbf{k}}) = 0$ except for $k_2 = 0$, their arguments can immediately be transfered to show that the weights $q^\beta(\underline{k})$ of the $\varepsilon(t)$ of (5b) satisfy (C1)-(C5) with $d = 1$ and $\Psi_1(\underline{n}) = const \cdot n_1$, $\Psi_{11}(\underline{n}) = const \cdot n_1^2$, $\Psi_{111}(\underline{n}) = const \cdot n_1^3$.

Let $\underline{e}_1 = \binom{1}{0}$, $\underline{e}_2 = \binom{0}{1}$, $\underline{1} = \binom{1}{1}$. As $a_{11} = \alpha_1$, $a_{12} = \alpha_2$, $\alpha_{21} = 1$ and $a_{22} = 0$, we have from Lemma A1 c) of the appendix for $P(\underline{m}, \underline{n}) \neq 0$

$$\frac{\partial}{\partial \alpha_1} \log P(\underline{m}, \underline{n}) = \frac{m_1}{1-\alpha_1} \left\{ \frac{P(\underline{m} - \underline{e}_1, \underline{n} - \underline{1})}{P(\underline{m}, \underline{n})} - 1 \right\} =: Q_1(\underline{m}, \underline{n})$$

$$\frac{\partial}{\partial \alpha_2} \log P(\underline{m}, \underline{n}) = \frac{m_2}{1-\alpha_2} \left\{ \frac{P(\underline{m} - \underline{e}_2, \underline{n} - \underline{e}_1)}{P(\underline{m}, \underline{n})} - 1 \right\} =: Q_2(\underline{m}, \underline{n})$$

Using the convention $q^\beta(-\underline{e}_1) = 0$, we have

$$\begin{aligned} \frac{\partial}{\partial \beta} q^\beta(\underline{k}) &= q^\beta(\underline{k} - \underline{e}_1) - q^\beta(\underline{k}) \quad \text{for} \quad k_1 \geq 0,\ k_2 = 0\,, \\ \frac{\partial}{\partial \beta} q^\beta(\underline{k}) &= 0 \qquad\qquad\qquad\qquad\quad \text{else.} \end{aligned}$$

Applying the explicit formula (A2) of the appendix for $P(\underline{m}, \underline{n})$, a straightforward calculation shows for $P(\underline{m}, \underline{n}) > 0$

$$\frac{\partial}{\partial \beta} \log P(\underline{m}, \underline{n}) = \frac{P(\underline{m}, \underline{n} - \underline{e}_1)}{P(\underline{m}, \underline{n})} - 1 =: Q_3(\underline{m}, \underline{n})$$

We can write the Fisher information matrix $\Sigma(\vartheta) = \Sigma(\alpha_1, \alpha_2, \beta)$ in terms of the functions Q_1, Q_2, Q_3, and, in particular, we have for $\underline{z} \in \mathbb{R}^3$

$$\underline{z}^T \Sigma(\vartheta) \underline{z} = \mathsf{E}\{\sum_{i=1}^{3} z_i Q_i(X(1), X(2)\}^2$$

as, of course, $P(X(1), X(2)) > 0$ a.s. with respect to the stationary distribution μ of $\{X(t)\}$. Therefore, $\Sigma(\vartheta)$ is positive definite and satisfies (C6) if

$$\sum_{i=1}^{3} z_i Q_i(X(1), X(2) = 0 \qquad \mu - a.s. \tag{6}$$

implies $\underline{z} = 0$. As the support of μ is $\mathbb{N}_0^2$ and as we have Poisson innovations,

$$P(\underline{m}, \underline{n}) > 0 \quad \text{for all} \quad \underline{m}, \underline{n} \in \mathbb{N}_0^2 \quad \text{with} \quad \mathrm{n}_2 = \mathrm{m}_1.$$

Therefore, (6) implies

$$\sum_{i=1}^{3} z_i Q_i \left(\binom{m_1}{m_2}, \binom{k}{m_1} \right) = 0 \qquad \text{for all} \quad m_1, m_2, k \geq 0\,.$$

Now, the special selections $m_1 = m_2 = 0$, $m_1 = 1$ and $m_2 = 0$, $m_1 = 0$ and $m_2 = 1$ immediately imply $z_3 = 0$, $z_1 = 0$ and $z_2 = 0$.

Appendix: Proof of Theorem 2

The proof of Theorem 2 is similar to the univariate case treated by Franke and Seligmann (1993). It is an application of a theorem of Billingsley (1961), and the basic idea for showing the assumptions of that general result is to use some recursive relations for the transition probabilites $P_\vartheta(\underline{\mathbf{m}}, \underline{\mathbf{n}})$ of the INAR(M)-process. To keep the somewhat tedious notation as simple as possible we restrict ourselves to the two-dimensional case $M = 2$, but it is easy to see from the following proof how the arguments have to be transferred to the case of general M. Furthermore, we just write $P(\underline{\mathbf{m}}, \underline{\mathbf{n}})$ instead of $P_\vartheta(\underline{\mathbf{m}}, \underline{\mathbf{n}})$ as the dependence of the transition probabilities on the parameter $\vartheta = (\alpha, \beta)$ is now obvious. The basic recursions are contained in the following Lemma.

<u>**Lemma A1:**</u> *Let* $\underline{\mathbf{e}}_1 = \binom{1}{0}$, $\underline{\mathbf{e}}_2 = \binom{0}{1}$, $\underline{\mathbf{1}} = \binom{1}{1}$.

a) $P(\underline{\mathbf{0}}, \underline{\mathbf{n}}) = q^\beta(\underline{\mathbf{n}})$ *for* $\underline{\mathbf{n}} \geq \underline{\mathbf{0}}$.
For $i = 1,\ j = 2$ *and vice versa*

$$\begin{aligned} P(\underline{\mathbf{m}}, \underline{\mathbf{n}}) &= a_{ii}\{a_{ji}P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_i, \underline{\mathbf{n}} - \underline{\mathbf{1}}) + (1 - a_{ji})P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_i, \underline{\mathbf{n}} - \underline{\mathbf{e}}_i)\} \\ &+ (1 - a_{ii})\{a_{ji}P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_i, \underline{\mathbf{n}} - \underline{\mathbf{e}}_j) + (1 - a_{ji})P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_i, \underline{\mathbf{n}})\} \end{aligned}$$

if $m_i \geq 1,\ m_j \geq 0$, $\underline{\mathbf{n}} \geq \underline{\mathbf{0}}$.

b) *For* $i = 1,\ j = 2$ *and vice versa let for* $\underline{\mathbf{m}} \geq \underline{\mathbf{0}}$, $h \geq 0$

$$\begin{aligned} P_i(\underline{\mathbf{m}}, h) &= pr(a_{ii} \circ m_i + a_{ij} \circ m_j = h) \\ &= \sum_{\mu=(h-m_j)^+}^{m_i \wedge h} \binom{m_i}{\mu}\binom{m_j}{h-\mu} a_{ii}^{\mu} a_{ij}^{h-\mu}(1 - a_{ii})^{m_i-\mu}(1 - a_{ij})^{m_j-h+\mu} \end{aligned} \tag{A1}$$

such that we can write

$$P(\underline{\mathbf{m}}, \underline{\mathbf{n}}) = \sum_{\underline{\mathbf{k}} \leq \underline{\mathbf{n}}} P_1(\underline{\mathbf{m}}, n_1 - k_1)\, P_2(\underline{\mathbf{m}}, n_2 - k_2)\, q^\beta(\underline{\mathbf{k}}). \tag{A2}$$

Then, for $i = 1, j = 2$ *and vice versa,* $m_j \geq 1, m_i \geq 0, h \geq 0$:

$$P_i(\underline{\mathbf{m}}, h) = a_{ij}\, P_i(\underline{\mathbf{m}} - \underline{\mathbf{e}}_j, h - 1) + (1 - a_{ij})\, P_i(\underline{\mathbf{m}} - \underline{\mathbf{e}}_j, h),$$

where, for convenience, we set $P_i(\underline{\mathbf{m}}, h) = 0$ *for* $h < 0$.

c) For $i = 1,\ j = 2$ and vice versa, $\underline{\mathbf{m}}, \underline{\mathbf{n}} \geq 0$:

$$\frac{\partial}{\partial a_{ii}} P(\underline{\mathbf{m}}, \underline{\mathbf{n}}) = \frac{m_i}{1-a_{ii}}\{a_{ji} P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_i, \underline{\mathbf{n}} - \underline{\mathbf{1}}) + (1 - a_{ji}) P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_i, \underline{\mathbf{n}} - \underline{\mathbf{e}}_i) - P(\underline{\mathbf{m}}, \underline{\mathbf{n}})\},$$

if $0 < a_{ii} < 1$.

$$\frac{\partial}{\partial a_{ij}} P(\underline{\mathbf{m}}, \underline{\mathbf{n}}) = \frac{m_j}{1-a_{ij}}\{a_{jj} P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_j, \underline{\mathbf{n}} - \underline{\mathbf{1}}) + (1 - a_{jj}) P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_j, \underline{\mathbf{n}} - \underline{\mathbf{e}}_i) - P(\underline{\mathbf{m}}, \underline{\mathbf{n}})\},$$

if $0 < a_{ij} < 1$.

Proof:

1. Let $a_{ij} \circ m_j = \sum_{\nu=1}^{m_j} Y_\nu^{(ij)}$, $i, j = 1, 2$,
with i.i.d. Bernoulli-variables $Y_1^{(ij)}, \ldots, Y_{m_j}^{(ij)}$ such that $a_{ij} = pr(Y_\nu^{(ij)} = 1)$. Therefore, for $m_j \geq 1$, we have
$$a_{ij} \circ m_j - Y_{m_j}^{(ij)} \stackrel{d}{=} a_{ij} \circ (m_j - 1).$$
Partitioning the event $\{A \circ \underline{\mathbf{m}} + \varepsilon(t) = \underline{\mathbf{n}}\}$ for $m_1 \geq 1$ w.r.t. the four cases $\{Y_{m_1}^{(11)} = 1, Y_{m_1}^{(21)} = 1\}$, $\{Y_{m_1}^{(11)} = 1, Y_{m_1}^{(21)} = 0\}$, $\{Y_{m_1}^{(11)} = 0, Y_{m_1}^{(21)} = 1\}$ $\{Y_{m_1}^{(11)} = 0, Y_{m_1}^{(21)} = 0\}$ and using independence, we get:
$$\begin{aligned} P(\underline{\mathbf{m}}, \underline{\mathbf{n}}) &= a_{11}a_{21}\, P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_1, \underline{\mathbf{n}} - \underline{\mathbf{1}}) + a_{11}(1 - a_{21})\, P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_1, \underline{\mathbf{n}} - \underline{\mathbf{e}}_1) \\ &+ (1 - a_{11})a_{21}\, P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_1, \underline{\mathbf{n}} - \underline{\mathbf{e}}_2) + (1 - a_{11})(1 - a_{21})\, P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_1, \underline{\mathbf{n}}) \end{aligned}$$
which is the recursion a) for $i = 1, j = 2$. By exchanging the indices 1 and 2, we get the other relation with $i = 2, j = 1$, too. The same type of argument also provides the recursion b).

2. If $m_i = 0$, $P(\underline{\mathbf{m}}, \underline{\mathbf{n}})$ does not depend on a_{1i} and a_{2i}, and the relations c) hold. Therefore, we only have to discuss the case $m_i \geq 1$, and we restrict ourselves to the situation $i = 1, j = 1$ as the other selections of indices can be dealt with analogously. We have
$$\begin{aligned} \frac{\partial}{\partial a_{11}} P_1(\underline{\mathbf{m}}, h) &= \sum_{\mu=(h-m_2)^+}^{m_1 \wedge h} \binom{m_1}{\mu}\binom{m_2}{h-\mu} \Big\{\mu\, a_{11}^{\mu-1}(1 - a_{11})^{m_1-\mu} \\ &\quad -(m_1 - \mu)a_{11}^{\mu}(1 - a_{11})^{m_1-\mu-1}\Big\} \cdot a_{12}^{h-\mu}(1 - a_{12})^{m_2-h+\mu} \\ &= -\frac{m_1}{1 - a_{11}} P_1(\underline{\mathbf{m}}, h) + \sum_{\mu=(h-m_2)\vee 1}^{m_1 \wedge h} \mu \binom{m_1}{\mu}\binom{m_2}{h-\mu} a_{11}^{\mu-1}(1 - a_{11})^{m_1-\mu-1} \\ &\quad a_{12}^{h-\mu}(1 - a_{12})^{m_2-h+\mu} \end{aligned}$$

$$= \frac{m_1}{1-a_{11}} \Bigg\{ -P_1(\underline{m}, h) + \sum_{\mu=(h-m_2)\vee 1}^{m_1 \wedge h} \binom{m_1 - 1}{\mu - 1} \binom{m_2}{h-\mu} a_{11}^{\mu-1} (1-a_{11})^{m_1-\mu}$$
$$a_{12}^{h-\mu}(1-a_{12})^{m_2-h+\mu}$$
$$= \frac{m_1}{1-a_{11}} \Big\{ P_1(\underline{m} - \underline{e}_1, h-1) - P_1(\underline{m}, h) \Big\}$$

Using this relation and (A2), we get by applying the recursion b) for $i = 2$, $j = 1$

$$\begin{aligned}
\frac{\partial}{\partial a_{11}} P(\underline{m}, \underline{n}) &= \frac{m_1}{1-a_{11}} \Bigg\{ \sum_{\underline{k} \le \underline{n}} P_1(\underline{m} - \underline{e}_1, n_1 - k_1 - 1)\, P_2(\underline{m}, n_2 - k_2)\, q^\beta(\underline{k}) \; - \; P(\underline{m}, \underline{n}) \Bigg\} \\
&= \frac{m_1}{1-a_{11}} \Bigg\{ a_{21} \sum_{\underline{k} \le \underline{n}} P_1(\underline{m} - \underline{e}_1, n_1 - k_1 - 1) \cdot P_2(\underline{m} - \underline{e}_1, n_2 - k_2 - 1)\, q^\beta(\underline{k}) \\
&\quad + (1 - a_{21}) \sum_{\underline{k} \le \underline{n}} P_1(\underline{m} - \underline{e}_1, n_1 - k_1 - 1)\, P_2(\underline{m} - \underline{e}_1, n_2 - k_2)\, q^\beta(\underline{k}) - P(\underline{m}, \underline{n}) \Bigg\} \\
&= \frac{m_1}{1-a_{11}} \Big\{ a_{21}\, P(\underline{m} - \underline{e}_1, \underline{n} - \underline{1}) + (1 - a_{21})\, P(\underline{m} - \underline{e}_1, \underline{n} - \underline{e}_1) - P(\underline{m}, \underline{n}) \Big\}
\end{aligned}$$

■

<u>**Lemma A2:**</u>*For* $\underline{m}, \underline{n} \ge 0$, $i, j = 1, 2$, *we have for* $P(\underline{m}, \underline{n}) > 0$

$$-\frac{m_j}{1 - a_{ij}} \le \frac{\partial}{\partial a_{ij}} \log P(\underline{m}, \underline{n}) \le \frac{m_j}{a_{ij}} \quad \textit{if} \quad 0 < a_{ij} < 1.$$

Proof: We only discuss the case $i = j = 1$, as the other situations can be treated analogously.
Using the abbreviations

$$S_{21} = a_{21}\, P(\underline{m} - \underline{e}_1, \underline{n} - \underline{1}) + (1 - a_{21})\, P(\underline{m} - \underline{e}_1, \underline{n} - \underline{e}_1) \ge 0,$$
$$S_{21}^* = a_{21}\, P(\underline{m} - \underline{e}_1, \underline{n} - \underline{e}_2) + (1 - a_{21})\, P(\underline{m} - \underline{e}_1, \underline{n}) \ge 0,$$

we have by Lemma A1 for $m_1 \ge 1$, $m_2 \ge 0$, $\underline{n} \ge 0$

$$\begin{aligned}
P(\underline{m}, \underline{n}) &= a_{11} S_{21} + (1 - a_{11}) S_{21}^* \\
\frac{\partial}{\partial a_{11}} P(\underline{m}, \underline{n}) &= \frac{m_1}{1 - a_{11}} \{ S_{21} - P(\underline{m}, \underline{n}) \} \\
&= \frac{m_1}{a_{11}} \{ P(\underline{m}, \underline{n}) - S_{21}^* \} \\
&\le \frac{m_1}{a_{11}} P(\underline{m}, \underline{n})
\end{aligned}$$

On the other hand, again by Lemma A1,

$$\frac{\partial}{\partial a_{11}} P(\underline{\mathbf{m}}, \underline{\mathbf{n}}) \geq -\frac{m_1}{1-a_{11}} P(\underline{\mathbf{m}}, \underline{\mathbf{n}})$$

Both inequalities also hold for $m_1 = 0$ as, then, $\frac{\partial}{\partial a_{11}} P(\underline{\mathbf{m}}, \underline{\mathbf{n}}) = 0$. ■

__Lemma A3:__ *For $i = 1$, $j = 2$ and vice versa, we have for $P(\underline{\mathbf{m}}, \underline{\mathbf{n}}) > 0$*

$$\begin{aligned}
a_{ji} \frac{P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_i, \underline{\mathbf{n}} - \underline{\mathbf{1}})}{P(\underline{\mathbf{m}}, \underline{\mathbf{n}})} &\leq \frac{1}{a_{ii}} \quad \textit{if} \quad 0 < a_{ii} < 1, \\
(1 - a_{ji}) \frac{P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_i, \underline{\mathbf{n}} - \underline{\mathbf{e}}_i)}{P(\underline{\mathbf{m}}, \underline{\mathbf{n}})} &\leq \frac{1}{a_{ii}} \quad \textit{if} \quad 0 < a_{ii} < 1, \\
a_{jj} \frac{P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_j, \underline{\mathbf{n}} - \underline{\mathbf{1}})}{P(\underline{\mathbf{m}}, \underline{\mathbf{n}})} &\leq \frac{1}{a_{ij}} \quad \textit{if} \quad 0 < a_{ij} < 1, \\
(1 - a_{jj}) \frac{P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_j, \underline{\mathbf{n}} - \underline{\mathbf{e}}_i)}{P(\underline{\mathbf{m}}, \underline{\mathbf{n}})} &\leq \frac{1}{a_{ij}} \quad \textit{if} \quad 0 < a_{ij} < 1.
\end{aligned}$$

Proof: If $m_i \geq 1$, we have from Lemma A1c

$$\begin{aligned}
0 &\leq a_{ji} \frac{P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_i, \underline{\mathbf{n}} - \underline{\mathbf{1}})}{P(\underline{\mathbf{m}}, \underline{\mathbf{n}})} + (1 - a_{ji}) \frac{P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_i, \underline{\mathbf{n}} - \underline{\mathbf{e}}_i)}{P(\underline{\mathbf{m}}, \underline{\mathbf{n}})} \\
&= \frac{1 - a_{ii}}{m_i} \frac{\partial}{\partial a_{ii}} \log P(\underline{\mathbf{m}}, \underline{\mathbf{n}}) + 1 \\
&\leq \frac{1 - a_{ii}}{a_{ii}} + 1 = \frac{1}{a_{ii}}
\end{aligned}$$

by Lemma A2, such that we have the first and second inequality. For $m_i = 0$, they are satisfied trivially, as then by definition the left-hand sides vanish. The third and fourth inequality can be shown analogously. ■

__Proof of Theorem 2:__ The Theorem is a special case of Theorem 2.2 of Billingsley (1961). We only have to check that conditions (C1) - (C6) imply the conditions of this general result. We first remark that (C2) implies $\mathsf{E}||X(t)||^3 < \infty$ for the stationary solution of (2). This can be shown completely analogous to the proof of Theorem 2.1 of Du and Li (1991) where, among other things, the existence of the second moment of a univariate INAR-process $\{X(t)\}$ is concluded from $\mathsf{E}\,\varepsilon(t)^2 < \infty$.

1. Using the explicit representation for $P(\underline{\mathbf{m}}, \underline{\mathbf{n}})$ provided by (A1) and (A2) and condition (C3) we conclude that $P(\underline{\mathbf{m}}, \underline{\mathbf{n}})$ is three times continuously differentiable w.r.t. $\alpha_1, \ldots, \alpha_s$ and $\beta_1, \ldots, \beta_d$, noting that $\alpha_k = a_{i(k),j(k)} < 1$ by assumption. Furthermore, from $0 < \alpha_k = a_{i(k),j(k)} < 1$, $k = 1, \ldots, s$, and (C1) we have that for any

$\underline{\mathbf{m}}$ the set $\{\underline{\mathbf{n}}; P(\underline{\mathbf{m}},\underline{\mathbf{n}}) > 0\}$ does not depend on α and β. Therefore, $\log P(\underline{\mathbf{m}},\underline{\mathbf{n}})$ is well-defined except for a set of $P(\underline{\mathbf{m}},.)$-measure 0 which does not depend on the parameter values.

2. As $A \circ \underline{\mathbf{m}} \le (m_1+m_2)\underline{\mathbf{1}}$ with $\underline{\mathbf{1}} = \binom{1}{1}$, we have for $n_1, n_2 \ge m_1+m_2$

$$\begin{aligned} P(\underline{\mathbf{m}},\underline{\mathbf{n}}) &= \sum_{k_1=n_1-m_1-m_2}^{n_1} \sum_{k_2=n_2-m_1-m_2}^{n_2} pr(A\circ\underline{\mathbf{m}} = \underline{\mathbf{n}}-\underline{\mathbf{k}})\, q^\beta(\underline{\mathbf{k}}) \\ &\le \sum_{\underline{\mathbf{n}}-(m_1+m_2)\underline{\mathbf{1}}\le\underline{\mathbf{k}}\le\underline{\mathbf{n}}} q^\beta(\underline{\mathbf{k}}) \end{aligned}$$

The first relation of (C4), therefore, implies that for each ϑ' there exists a neighbourhood V such that for fixed $\underline{\mathbf{m}}$
$\sum_{\underline{\mathbf{n}}\ge\underline{\mathbf{0}}} \sup_{\vartheta\in V} P(\underline{\mathbf{m}},\underline{\mathbf{n}}) \le (m_1+m_2+1)^2\{1+\sum_{\underline{\mathbf{k}}\ge\underline{\mathbf{0}}} \sup_{\beta\in U} q^\beta(\underline{\mathbf{k}})\} < \infty,$

where U denotes the projection of V onto the subspace corresponding to parameters $\beta_1,\ldots,\beta_d$. By Lemma A1 c), the same summability condition holds for $\frac{\partial}{\partial\alpha_k}P(\underline{\mathbf{m}},\underline{\mathbf{n}})$ and $\frac{\partial^2}{\partial\alpha_k\partial\alpha_l}P(\underline{\mathbf{m}},\underline{\mathbf{n}})$, $k,l = 1,\ldots,s$, instead of $P(\underline{\mathbf{m}},\underline{\mathbf{n}})$. Using additionally the second and third relation of (C4), we have the uniform summability for all first and second derivatives of $P(\underline{\mathbf{m}},\underline{\mathbf{n}})$ w.r.t. parameters $\alpha_1,\ldots,\alpha_s$, $\beta_1,\ldots,\beta_d$,.

3. Calculating expectations w.r.t. the stationary state of $\{X(t)\}$ we have for $\alpha_k = a_{i(k),j(k)}$, using Lemma A2,

$$\begin{aligned} \mathsf{E}\left|\frac{\partial}{\partial\alpha_k}\log P(X(1),X(2))\right|^2 &\le C^2\cdot\mathsf{E}\{\max(X_{i(k)}(1),X_{j(k)}(1)\}^2 \\ &\le C^2\cdot\mathsf{E}(X_{i(k)}(1)+X_{j(k)}(1))^2 < \infty \end{aligned}$$

with $C = \max(\alpha_k^{-1},(1-\alpha_k)^{-1})$. Similarly, we have from (C5)

$$\text{(A3)}\left|\frac{\partial}{\partial\beta_u}\log P(\underline{\mathbf{m}},\underline{\mathbf{n}})\right| \le \frac{1}{P(\underline{\mathbf{m}},\underline{\mathbf{n}})}\sum_{\underline{\mathbf{k}}\le\underline{\mathbf{n}}} pr(A\circ\underline{\mathbf{m}}=\underline{\mathbf{n}}-\underline{\mathbf{k}})\left|\frac{\partial}{\partial\beta_u}q^\beta(\underline{\mathbf{k}})\right| \le \Psi_u(\underline{\mathbf{n}})$$

and $\mathsf{E}\left|\frac{\partial}{\partial\beta_u}\log P(X(1),X(2))\right|^2 \le \mathsf{E}\,\Psi_u^2(X(2)) < \infty$

Therefore, the Fisher information matrix $\Sigma(\vartheta)$ is well-defined, and, by (C6), it is nonsingular.

4. For all $\alpha_k = a_{i(k),j(k)}$, $k = 1,\ldots,s$, we have

$$\text{(A4)}\qquad \left|\frac{\partial}{\partial\alpha_k}\frac{\partial}{\partial\alpha_l}\log P(\underline{\mathbf{m}},\underline{\mathbf{n}})\right| \le C\cdot m_{j(k)}\, m_{j(l)},$$

where the constant C can be chosen uniform over a suitable neighbourhood of any $\vartheta' \in \Theta$. To show this relation one has to calculate the second derivatives explicitly using Lemma A1c, and then Lemma A2 and Lemma A3 are applied, e.g. for $k = l$ and $i(k) = j(k) = 1$ we have

$$\begin{aligned}
\frac{\partial^2}{\partial \alpha_k^2} \log P(\underline{\mathrm{m}}, \underline{\mathrm{n}}) &= \frac{\partial^2}{\partial a_{11}^2} \log P(\underline{\mathrm{m}}, \underline{\mathrm{n}}) \\
&= \frac{m_1}{(1-a_{11})^2} \{Q_{21} - 1\} + \frac{m_1}{1-a_{11}} \frac{\partial}{\partial a_{11}} \{Q_{21} - 1\}
\end{aligned}$$

with

$$Q_{21} = a_{21} \frac{P(\underline{\mathrm{m}} - \underline{\mathrm{e}}_1, \underline{\mathrm{n}} - \underline{1})}{P(\underline{\mathrm{m}}, \underline{\mathrm{n}})} + (1 - a_{21}) \frac{P(\underline{\mathrm{m}} - \underline{\mathrm{e}}_1, \underline{\mathrm{n}} - \underline{\mathrm{e}}_1)}{P(\underline{\mathrm{m}}, \underline{\mathrm{n}})}.$$

Q_{21} is bounded by Lemma A3. Using

$$\text{(A5)} \qquad \left| \frac{\partial}{\partial a_{11}} \frac{P(\underline{\mathrm{k}}, \underline{\mathrm{l}})}{P(\underline{\mathrm{m}}, \underline{\mathrm{n}})} \right| = \frac{P(\underline{\mathrm{k}}, \underline{\mathrm{l}})}{P(\underline{\mathrm{m}}, \underline{\mathrm{n}})} \left| \frac{\partial}{\partial a_{11}} \{\log P(\underline{\mathrm{k}}, \underline{\mathrm{l}}) - \log P(\underline{\mathrm{m}}, \underline{\mathrm{n}})\} \right|$$

for $P(\underline{\mathrm{k}}, \underline{\mathrm{l}}) > 0$ and $\frac{\partial}{\partial a_{11}} \frac{P(\underline{\mathrm{k}}, \underline{\mathrm{l}})}{P(\underline{\mathrm{m}}, \underline{\mathrm{n}})} = 0$ else, we see from Lemma A3 and Lemma A2 that $\frac{\partial}{\partial a_{11}} Q_{21}$ is bounded by *const.* $\cdot\ m_1$.
Analogously, we have for $k, l, u = 1, \ldots, s$

$$\text{(A6)} \qquad \left| \frac{\partial}{\partial \alpha_k} \frac{\partial}{\partial \alpha_l} \frac{\partial}{\partial \alpha_u} \log P(\underline{\mathrm{m}}, \underline{\mathrm{n}}) \right| \le C \cdot m_{j(k)}\, m_{j(l)}\, m_{j(u)}\,,$$

for a suitable locally uniform constant C. We illustrate the argument with the case $k = l = u,\ \ i(k) = j(k) = 1$

$$\begin{aligned}
\frac{\partial^2}{\partial \alpha_k^3} \log P(\underline{\mathrm{m}}, \underline{\mathrm{n}}) &= \frac{\partial^3}{\partial a_{11}^3} \log P(\underline{\mathrm{m}}, \underline{\mathrm{n}}) \\
&= \frac{2m_1}{(1-a_1)^3} \{Q_{21} - 1\} + \frac{2m_1}{(1-a_{11})^2} \frac{\partial}{\partial a_{11}} Q_{21} + \frac{m_1}{1-a_{11}} \frac{\partial^2}{\partial a_{11}^2} Q_{21}
\end{aligned}$$

Again by Lemma A3 and (A5) the first two terms on the right-hand side are bounded by *const* $\cdot\ m_1^2$. By (A5), $\frac{\partial^2}{\partial a_{11}^2} Q_{21}$ is a linear combination of terms of the form

$$\frac{\partial}{\partial a_{11}} \left\{ \frac{P(\underline{\mathrm{k}}, \underline{\mathrm{l}})}{P(\underline{\mathrm{n}}, \underline{\mathrm{n}})} \right\} \left[\frac{\partial}{\partial a_{11}} \{\log P(\underline{\mathrm{k}}, \underline{\mathrm{l}}) - \log P(\underline{\mathrm{m}}, \underline{\mathrm{n}})\} \right] \qquad \text{and}$$

$$\frac{P(\underline{\mathrm{k}}, \underline{\mathrm{l}})}{P(\underline{\mathrm{m}}, \underline{\mathrm{n}})} \left[\frac{\partial^2}{\partial a_{11}^2} \{\log P(\underline{\mathrm{k}}, \underline{\mathrm{l}}) - \log P(\underline{\mathrm{m}}, \underline{\mathrm{n}})\} \right],$$

with $(\underline{\mathbf{k}}, \underline{\mathbf{l}}) \in \{(\underline{\mathbf{m}} - \underline{\mathbf{e}}_1, \underline{\mathbf{n}} - \underline{\mathbf{1}}), (\underline{\mathbf{m}} - \underline{\mathbf{e}}_1, \underline{\mathbf{n}} - \underline{\mathbf{e}}_1)\}$ and coefficients a_{21} or $(1 - a_{21})$, such that $\frac{\partial^2}{\partial a_{11}^2} Q_{21}$ is bounded by $const \cdot m_1^2$ too, using Lemma A3, Lemma A2 and (A4).

5. We have to show that local suprema of all third order derivatives of $\log P(X_1, X_2)$ have a finite mean. For this purpose, we get locally uniform bounds on all third derivatives of $\log P(\underline{\mathbf{m}}, \underline{\mathbf{n}})$ by using the same arguments as in 4. and the inequalities of condition (C5). We have, e.g., from Lemma A1c

$$\begin{aligned}\frac{\partial}{\partial \beta_u} \frac{\partial}{\partial a_{11}} \log P(\underline{\mathbf{m}}, \underline{\mathbf{n}}) &= \frac{m_1}{1 - a_{11}} \left\{ a_{21} \frac{\partial}{\partial \beta_u} \frac{P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_1, \underline{\mathbf{n}} - \underline{\mathbf{1}})}{P(\underline{\mathbf{m}}, \underline{\mathbf{n}})} + (1 - a_{21}) \frac{\partial}{\partial \beta_u} \frac{P(\underline{\mathbf{m}} - \underline{\mathbf{e}}_1, \underline{\mathbf{n}} - \underline{\mathbf{e}}_1)}{P(\underline{\mathbf{m}}, \underline{\mathbf{n}})} \right\}, \\ &= \frac{m_1}{1 - a_{11}} \frac{\partial}{\partial \beta_u} Q_{21}\end{aligned}$$

Using as in (A5)

$$\left| \frac{\partial}{\partial \beta_u} \frac{P(\underline{\mathbf{k}}, \underline{\mathbf{l}})}{P(\underline{\mathbf{m}}, \underline{\mathbf{n}})} \right| = \frac{P(\underline{\mathbf{k}}, \underline{\mathbf{l}})}{P(\underline{\mathbf{m}}, \underline{\mathbf{n}})} \left| \frac{\partial}{\partial \beta_u} \{\log P(\underline{\mathbf{k}}, \underline{\mathbf{l}}) - \log P(\underline{\mathbf{m}}, \underline{\mathbf{n}})\} \right|,$$

we see from Lemma A3 and (A3) that $\frac{\partial}{\partial \beta_u} Q_{21}$ is bounded by $const \cdot \Psi_u(\underline{\mathbf{n}})$. Therefore, we have for $\alpha_k = a_{11}$, $i(k) = j(k) = 1$

$$\frac{\partial}{\partial \beta_u} \frac{\partial}{\partial \alpha_k} \log P(\underline{\mathbf{m}}, \underline{\mathbf{n}}) \leq const\ m_{j(k)} \Psi_u(\underline{\mathbf{n}}),$$

and this relation holds for all $u = 1, \ldots, d$, $k = 1, \ldots, s$. Similarly, for $\alpha_k = a_{11}$

$$\frac{\partial}{\partial \beta_u} \frac{\partial^2}{\partial \alpha_k} \log P(\underline{\mathbf{m}}, \underline{\mathbf{n}}) = \frac{m_1}{(1 - a_{11})^2} \frac{\partial}{\partial \beta_u} Q_{21} + \frac{m_1}{1 - a_{11}} \frac{\partial}{\partial \beta_u} \frac{\partial}{\partial a_{11}} Q_{21}$$

is bounded by $const\ m_1^2 \Psi_u(\underline{\mathbf{n}})$, as the first term on the right-hand side is bounded by $m_1 \Psi_u(\underline{\mathbf{n}})$, and by (A5), $\frac{\partial}{\partial \beta_u} \frac{\partial}{\partial a_{11}} Q_{21}$ is a linear combination of terms of the form

$$\frac{\partial}{\partial \beta_u} \left\{ \frac{P(\underline{\mathbf{k}}, \underline{\mathbf{l}})}{P(\underline{\mathbf{m}}, \underline{\mathbf{n}})} \right\} \left[\frac{\partial}{\partial a_{11}} \{\log P(\underline{\mathbf{k}}, \underline{\mathbf{l}}) - \log P(\underline{\mathbf{m}}, \underline{\mathbf{n}})\} \right]$$

and

$$\frac{P(\underline{\mathbf{k}}, \underline{\mathbf{l}})}{P(\underline{\mathbf{m}}, \underline{\mathbf{n}})} \left[\frac{\partial}{\partial \beta_u} \frac{\partial}{\partial a_{11}} \{\log P(\underline{\mathbf{k}}, \underline{\mathbf{l}}) - \log P(\underline{\mathbf{m}}, \underline{\mathbf{n}})\} \right]$$

with $(\underline{\mathbf{k}}, \underline{\mathbf{l}}) \in \{(\underline{\mathbf{m}} - \underline{\mathbf{e}}_1, \underline{\mathbf{n}} - \underline{\mathbf{1}}), (\underline{\mathbf{m}} - \underline{\mathbf{e}}_1, \underline{\mathbf{n}} - \underline{\mathbf{e}}_1)\}$ and coefficients a_{21} or $(1 - a_{21})$, such that the second term is bounded by $const\ m_1^2 \Psi_u(\underline{\mathbf{n}})$, using Lemma A2 and the bound on $\frac{\partial}{\partial \beta_u} \frac{\partial}{\partial a_{11}} \log P(\underline{\mathbf{m}}, \underline{\mathbf{n}})$ from above.

Using similar arguments, we have additionally to (A6)

$$\begin{aligned}
\left|\frac{\partial}{\partial\beta_u}\frac{\partial}{\partial\alpha_k}\frac{\partial}{\partial\alpha_l}\log P(\underline{\mathbf{m}},\underline{\mathbf{n}})\right| &\le C\, m_{j(k)}\, m_{j(l)}\, \Psi_u(\underline{\mathbf{n}}) \\
\left|\frac{\partial}{\partial\beta_u}\frac{\partial}{\partial\beta_v}\frac{\partial}{\partial\alpha_k}\log P(\underline{\mathbf{m}},\underline{\mathbf{n}})\right| &\le C\, m_{j(k)}\, \{\Psi_{uv}(\underline{\mathbf{n}}) + \Psi_u(\underline{\mathbf{n}})\,\Psi_v(\underline{\mathbf{n}})\} \\
\left|\frac{\partial}{\partial\beta_u}\frac{\partial}{\partial\beta_v}\frac{\partial}{\partial\beta_w}\log P(\underline{\mathbf{m}},\underline{\mathbf{n}})\right| &\le \Psi_{uvw}(\underline{\mathbf{n}}) + \Psi_u(\underline{\mathbf{n}})\,\Psi_{vw}(\underline{\mathbf{n}}) + \Psi_v(\underline{\mathbf{n}})\,\Psi_{uw}(\underline{\mathbf{n}}) \\
&\quad + \Psi_w(\underline{\mathbf{n}})\,\Psi_{uv}(\underline{\mathbf{n}}) + 2\Psi_u(\underline{\mathbf{n}})\,\Psi_v(\underline{\mathbf{n}})\,\Psi_w(\underline{\mathbf{n}})
\end{aligned}$$

for all $k, l = 1, \ldots, s$ and $u, v, w = 1, \ldots, d$. Using (C5) and $\mathsf{E}||X_1||^3 < \infty$, we finally have for all parameter values $\vartheta' \in \Theta$ the existence of a neighbourhood V such that for all third order derivatives

$$\mathsf{E}\sup_{\vartheta\in V}\left|\frac{\partial}{\partial\vartheta_u}\frac{\partial}{\partial\vartheta_v}\frac{\partial}{\partial\vartheta_w}\log P(X_1, X_2)\right| < \infty$$

References

[1] Al-Osh, M.A., and Alzaid, A.A. (1987). First order integer-valued autoregressive (INAR(1)-) processes. *J. Time Ser. Anal.* **8**, 261-275.

[2] Basilevsky, A. (1983). *Applied Matrix Algebra in the Statistical Sciences.* North Holland, New York-Amsterdam-Oxford.

[3] Billingsley, P. (1961). *Statistical Inference for Markov processes.* University of Chicago Press, Chicago.

[4] Du Jin-Guan, and Li Yuan (1991). The integer-valued autoregressive (INAR(p)-) model. *J. Time Ser. Anal.* **12**, 129-142.

[5] Franke, J., and Seligmann, Th. (1993). *Conditional maximum-likelihood estimates for INAR(1)-processes and their application to modelling epileptic seizure counts.* In: Developments in Time Series Analysis, ed. T. Subba Rao, Chapmann & Hall, London.

[6] McKenzie, E. (1985). Discussion of Modelling and Residual Analysis of Nonlinear Autoregressive Time Series in Exponential Variables by A.J. Lawrence and P.A.W. Lewis. *JRSS Ser. B.* **47**, 187-188.

[7] McKenzie, E. (1986). ARMA processes with negative binomial and geometric marginal distributions. *Adv. Appl. Prob.*, **18**, 679-705.

[8] McKenzie, E. (1987). Innovation distributions for Gamma and negative binomial autoregressions. *Scand. J. Statist.*, **14**, 79-85.

[9] McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Adv. Appl. Prob.* **20**, 822-835.

[10] Rosenblatt, M. (1971). *Markov Processes. Structure and Asymptotic Behaviour.* Springer, Berlin-Heidelberg-New York.

[11] Rosenblatt, M. (1974). *Random processes, 2nd edition.* Springer, Berlin-Heidelberg-New York.

[12] Seneta, E. (1969). Functional equations and the Galton-Watson process. *Adv. Appl. Prob.*, **1**, 1-42.

[13] Venkataraman, K.N. (1982) A time series approach to the study of the simple subcritical Galton-Watson process with immigration. *Adv. Appl. Prob.*, **14**, 1-20.

[14] Venkataraman, K.N. and Nanthi, K. (1982). A limit theorem on subcritical Galton-Watson process with immigration. *Ann. Probab.*, **10**, 1069-1074.