Local asymptotic normality and efficient estimation for multivariate GINAR(p) models

Hiroshi Shiraishi*

Abstract: We derive the Local Asymptotic Normality (LAN) property for a multivariate generalized integer-valued autoregressive (MGINAR) process with order p. The generalized thinning operator in the MGINAR(p) process includes not only the usual Binomial thinning but also Poisson thinning, geometric thinning, Negative Binomial thinning and so on. By using the LAN property, we propose an efficient estimation method for the parameter of the MGINAR(p) process. Our procedure is based on the one-step method, which update initial $\sqrt{n}$-consistent estimators to efficient ones. The one-step method has advantages in both computational simplicity and efficiency. Some numerical results for the asymptotic relative efficiency (ARE) of our estimators and the CLS estimators are presented. In addition, a real data analysis is provided to illustrate the application of the proposed estimation method.

Keywords: integer-valued time series; thinning operations; LAN; efficient estimation

1. Introduction

Recently, there has been a growing interest in modelling discrete-valued time series that arise in various fields of statistics (e.g., Weiß, 2018). This paper is concerned with a special class of observation-driven models termed “integer-valued autoregressive processes” (INAR processes), which were introduced independently by Al-Osh and Alzaid (1987) and McKenzie (1985, 1988). They introduced the
integer-valued autoregressive process with order 1 (INAR(1) process) to model non-negative integer-valued phenomena with time dependence. The more general INAR(p) processes were considered by Alzaid and Al-Osh (1990), Du and Li (1991) and so on. The INAR process consists of a mixture both of the distribution of thinning operator and the distribution of the innovation process. Alzaid and Al-Osh (1990) discussed the INAR(p) process in case that the thinning operator follows Binomial distribution (i.e., specified), while the distribution of the innovation process is left unspecified. Latour (1998) introduced a generalized version of the INAR(p) process, namely, the causal and stationary generalized integer-valued autoregressive (GINAR(p)) process. The generalized thinning operator in the GINAR(p) process includes not only the usual Binomial thinning but also Poisson thinning, Geometric thinning, and Negative Binomial thinning (by Ristić, Bakouch, & Nastić, 2009). The above history is appropriate for the univariate case. Nowadays, extensions for the multivariate case are being actively researched and applied. One of the first approaches to multivariate thinning mechanism was by McKenzie (1988). After that, Franke and Subba Rao (1993) introduced a multivariate INAR(1) (MINAR(1)) model based on independent Binomial thinning operators. The extensions for the MINAR(1) model were discussed by Latour (1997), Karlis and Pedeli (2013), and so on. This paper discusses a parameter estimation problem for a multivariate version of the GINAR(p) (MGINAR(p)) process. The MGINAR model is quite a large class, including all the models described above.

Estimation of the parameter for the INAR(p) process can be carried out in a variety of ways. Common ways for estimating parameters include the method of moments (MM), based on the Yule-Walker equations, and conditional least squares (CLS). The main advantage of both approaches is their simplicity due to the closed-form formulae and robustness due to requiring no assumption for the distribution. It is known that MM and CLS estimators are asymptotically equivalent. However, Al-Osh and Alzaid (1987) and so on have recommended using maximum likelihood (ML) estimators instead of MM and CLS estimators because they are less biased for small sample sizes. However, it is well known that the ML method is computationally unattractive due to complicated transition probabilities, including many convolutions. To overcome this problem, Drost, Van Den Akker, and Werker (2008) considered one-step, asymptotically efficient estimation of the INAR(p) model. Their method can reduce high computational cost due to the convolutions involved in the ML method. Following Drost et al. (2008), this paper provides a one-step method, which update initial $\sqrt{n}$ – consistent estimators to efficient ones for the MGINAR(p) processes. In the class of multivariate integer-valued models, the number of convolutions involved in the likelihood function is very large. For some distributions with reproductive property, the likelihood can be simplified, but it is impossible to apply such a simplification for all models in the MGINAR(p) processes. Therefore, it would be quite important to reduce the computational cost by using a one-step estimation. We first establish the local asymptotic normality (LAN) structure for experiments of the MGINAR(p) process. Considering the CLS estimator as the initial estimator, a one-step update estimator is proposed. We can also show that this estimator is asymptotically efficient.

The organization of this paper is as follows. In Section 2, the MGINAR(p) process is introduced and the LAN property is established. Section 3 discusses the efficient estimation. The CLS estimator is introduced and its asymptotic property is shown. Then, by using the LAN property, a one-step estimator is proposed to update the CLS estimator. In Section 4, the asymptotic relative efficiency (ARE) of the one-step estimator and the CLS estimator is examined through some simulation experiments. In addition, a real data analysis is provided to illustrate the application of the proposed estimation method. The proofs and other details are included in the Appendix.

2. The LAN property

Let $\{X_t = (X_{1,t}, \ldots, X_{d,t})^T; t \in \mathbb{Z}_+ = \mathbb{N} \cup \{0\}\}$ be a $d$- dimensional non-negative integer-valued random process (i.e., $X_{i}(\omega) \in \mathbb{Z}_+^d$). The multivariate generalized integer-valued autoregressive process of order $p$ (MGINAR(p)) is defined by
\[ X_t = \sum_{k=1}^{p} A^{(k)} \circ X_{t-k} + \epsilon_t \]  

(1)

where \( A^{(k)} = (A_{ij}^{(k)})_{i,j=1,...,d} \) is a \( d \times d \)-matrix for \( k = 1, \ldots, p \) and the matrix thinning \( A^{(k)} \circ X_{t-k} \) gives a \( d \)-dimensional random vector with \( t \)th component

\[ \left[ A^{(k)} \circ X_{t-k} \right]_i = \sum_{j=1}^{d} A^{(k)}_{ij} \circ X_{t-k}^j = \sum_{j=1}^{d} \sum_{r=1}^{d} \xi_{ijr}^{(k)}, \quad i = 1, \ldots, d. \]  

(2)

Note that for any \( A \in \mathbb{R}, A \circ 0 = 0 \); \( \{ \xi_{ijr}^{(k)}; r \in \mathbb{Z}_+ \} \) is a collection of independent and identically distributed (i.i.d.), non-negative, integer-valued random variables with the distribution function \( G_{A^{(k)}} \) and the mean \( A^{(k)}_j : \{ \epsilon_t = (\epsilon_1, \ldots, \epsilon_{d_T})^T; t \in \mathbb{Z}_+ \} \) is a collection of i.i.d. non-negative, integer-valued random vectors, where the \( j \)th component \( \epsilon_j \) has an independent distribution function \( F_{\epsilon_j} \) and the mean \( \omega_j \). Suppose that the starting values \( \{ X_0, \ldots, X_{-1} \} \) have a distribution \( \nu \) on \( \mathbb{Z}_{d_T}^d \), and \( \{ X_0, \ldots, X_{-1} \} \), \( \{ \xi_{ijr}^{(1)} \}, \ldots, \{ \xi_{ijr}^{(p)} \} \), \( \{ \epsilon \} \).

are independent of each other.

Throughout the paper, the number of logs \( (p \in \mathbb{N}) \) and the dimensions \( (d \in \mathbb{N}) \) are fixed and known. Let \( F \) and \( G \) denote the classes of distributions of \( \epsilon \) and \( \zeta \) that belong to parametric classes, respectively, say \( F = (F_0; \omega \in \Theta_0) \) and \( G = (G_0; A \in \Theta_0) \). For instance, we consider the Binomial thinning operator and the Binomial innovation, \( F \) and \( G \) should be defined by \( F_0(x) = a^n(1-a)^{n-x} \) and \( G_0(x) = A^n(1-A)^{n-x} \) for \( \omega \in \Theta_0 = (0,1), A \in \Theta_0 = (0,1) \), and \( x \in (0,1) \). Our goal is to estimate the parameter \( \theta = (a^T, A^T)^T \in \Theta \subset \mathbb{R}^q \) with \( q = d + d^2p \), efficiently, where

\[ \Theta = \left\{ (a^T, A^T)^T; a = (a_1, \ldots, a_d)^T \in \Theta^d_0; A = \text{vec}(A^{(1)}, \ldots, A^{(p)})^T = (A^{(1)}_1, \ldots, A^{(p)}_d)^T \in \Theta^{d^2p} \right\}. \]

For (probability) measures \( F_1, \ldots, F_k \in F \cup G \), we introduce the following notations for the convolutions:

- \( F_1 \bowtie F_2 \): the convolution of \( F_1 \) and \( F_2 \)
- \( \oplus_{k=1}^{n} F_k \): the convolution of \( F_1, \ldots, F_k \) (i.e., \( \oplus_{k=1}^{n} F_k = F_1 \bowtie \cdots \bowtie F_k \))
- \( F_1^{\circ n} \): the \( n \) times convolution of \( F_1 \) (i.e., \( F_1^{\circ n} = F_1 \bowtie \cdots \bowtie F_1 \))

Based on the above notations, we consider the corresponding probability space for \( \{ X_t \} \) denoted by \( (\Omega, \mathcal{F}, \mathbb{P}_{\nu,F,G}) \), where \( \Omega \) is a sample space, \( \mathcal{F} \) is the \( \sigma \)-algebra, and \( \mathbb{P}_{\nu,F,G} \) is the probability measure given by \( \nu \) (distribution of \( \{ X_{-p}, \ldots, X_{-1} \} \), \( \theta \) (parameter), \( F \) (class of distribution of \( \epsilon \)) and \( G \) (class of distribution of \( \zeta \)). Furthermore, we introduce \( F = \{ F_t; t \geq -p \} \) as the natural filtration generated by \( \{ X_{-p}, \ldots, X_t \} \) (i.e., \( F_t = \sigma(\{ X_{-p}, \ldots, X_t \}) \neq \sigma(\{ X_{-p}, \ldots, X_{-1}, \epsilon_{-p}, \ldots, \epsilon_t \}) \)).

From (1) and (2), we can write

\[ X_t = \sum_{k=1}^{p} \sum_{j=1}^{d} \sum_{r=1}^{d} \xi_{ijr}^{(k)} + \epsilon_t \]

where \( \xi_{ijr}^{(k)} = (\xi_{ijr}^{(1)}, \ldots, \xi_{ijr}^{(p)})^T \). Hence, it follows, for any \( t \in \mathbb{Z}_+ \),

\[ \mathbb{P}_{\nu,F,G}(X_t = x_t | F_{t-1}) = \mathbb{P}_{\nu,F,G}(X_t = x_t | X_{t-1}, \ldots, X_{-p}) \]

:= \( p_{\nu,F,G}(x_{t-1}, \ldots, x_{-p}, x_t) \) (say)

For some \( x_{t-k} = (x_{t-1-k}, \ldots, x_{t-d-k})^T \in \mathbb{Z}_{d_T}^d, k = 0, 1, \ldots, p \), the transition probability \( p_{\nu,F,G}^{(k)}(x_{t-1-k}, \ldots, x_{t-p}, x_t) \) is given by
\[
p_{\Theta}^{\alpha_1, \ldots, \alpha_d, \beta} = P_{\alpha, \beta, F, G}\{X_t = x_t|X_{t-1} = x_{t-1}, \ldots, X_{t-p} = x_{t-p}\}
\]
\[
= P_{\alpha, \beta, F, G}\left\{\prod_{k=1}^{d} \left(\sum_{j=1}^{p} \sum_{r=1}^{d} \sum_{i=1}^{\gamma_{ik}} \epsilon_{ik} \right) + \epsilon_{t} = x_{t}|X_{t-1} = x_{t-1}, \ldots, X_{t-p} = x_{t-p}\right\}
\]
\[
= \prod_{t=1}^{d} P_{\alpha, \beta, F, G}\left\{\sum_{k=1}^{d} \sum_{j=1}^{p} \sum_{r=1}^{d} \sum_{i=1}^{\gamma_{ik}} \epsilon_{ik} \right\} = F_{\alpha, \beta}\left(\sum_{k=1}^{d} \sum_{j=1}^{p} \sum_{r=1}^{d} \sum_{i=1}^{\gamma_{ik}} \epsilon_{ik}\right)
\]

We consider parametric MGINAR models in which the parameter space is restricted to the stationary parameter space, for instance, in the Binomial thinning operator and \(d = 1\), the thinning part of the parameter space should be defined by \(\{A; A = (A^{(1)}, \ldots, A^{(p)})^T \in (0, 1)^d, \sum_{k=1}^{d} A^{(k)} < 1\}\). Suppose that \(\mathcal{F} \times \mathcal{G}\) is a combination of a family of parametric distributions for the thinning operator and innovation (immigrant) with the formula below:

\[
\mathcal{F} = \{F_{\alpha} = (F_{\alpha_1}, \ldots, F_{\alpha_d}); \alpha = (\alpha_1, \ldots, \alpha_d)^T \in \mathbb{R}^d\}
\]
\[
\mathcal{G} = \{G_A = (G_{A^{(1)}}, \ldots, G_{A^{(d)}}); A = \text{vec}(A^{(1)}, \ldots, A^{(p)})^T = (A^{(1)}, \ldots, A^{(p)})^T \in \mathbb{R}^{d^2}\}
\]

Suppose that for any \(\Theta = (\alpha^T, A^T)^T \in \Theta\), \(\{X_t\}\) with \(F_{\alpha} \times G_A \in \mathcal{F} \times \mathcal{G}\) is a strictly stationary process. Let \(P_{\alpha, \beta, F, G}^{(n)}\) denote the law of \(\{X_{n}, \ldots, X_n\}\) on the measurable space \(\left(Z_1^{(n+1); p}, N^{(n+1); p}\right)\) under \(P_{\alpha, \beta, F, G}\). Here \(Z_1^{(n+1); p}\) is the sample space and the power set \(\mathbb{N}^{(n+1); p}\) is the \(\sigma\)-algebra on this sample space. Observing \(\{X_{1}, \ldots, X_n\}\) yields the following sequence of experiments:

\[
\mathcal{E}^{(n)}(\nu, \Theta, F, G) = \left(Z_1^{(n+1); p}, N^{(n+1); p}, \left(P_{\alpha, \beta, F, G}^{(n)}; \Theta \in \Theta\right)\right), \quad n \in \mathbb{Z}_+
\]

where the initial distribution \(\nu\) is fixed and the distributions for the thinning operator and innovation (immigrant) are parametrized (i.e., once \(\Theta = (\alpha^T, A^T)^T \in \Theta\) is fixed, the distributions are given by \(F_{\alpha}\) and \(G_A\)).

To prove the LAN property for the sequence of experiments \(\mathcal{E}^{(n)}(\nu, \Theta, F, G)\), \(n \in \mathbb{Z}_+\), we impose the following assumptions.

**Assumption 1.** \(\{A\} \Theta\) is an open, convex subset of \(\mathbb{R}^d\) with \(q = d + d^2\).

**Assumption 2.** The supports of \(F_\alpha\) and \(G_A\) do not depend on \(\alpha\) and \(A\) and we have, for all \(\Theta = (\alpha^T, A^T)^T\), \(F_\alpha(0) \times G_A(0) \in (0, 1)^d \times (0, 1)^{d^2}p\).

**Assumption 3.** For all \(e \in \mathbb{Z}_+\) and all \(\Theta = (\alpha^T, A^T)^T = ((\alpha_i), (A_j^{(k)}))^T \in \Theta\),

\[
f_{\alpha}(e) = \frac{\partial F_{\alpha}}{\partial \alpha_i}(e), \quad \hat{f}_{\alpha}(e) = \frac{\partial^2 F_{\alpha}}{\partial \alpha_i \partial A_j^{(k)}}(e), \quad g_{\alpha_i}(e) = \frac{\partial G_{A_j^{(k)}}}{\partial \alpha_i}(e), \quad \hat{g}_{\alpha_i}(e) = \frac{\partial^2 G_{A_j^{(k)}}}{\partial \alpha_i \partial A_j^{(k)}}(e)
\]

are defined, continuous in \(\Theta\), and satisfied with \(\frac{\partial^2 G_{A_j^{(k)}}}{\partial \alpha_i \partial A_j^{(k)}} = 0\) for \(i \neq j\), \((i, j, k) \neq (i', j', k')\), respectively.

**Assumption 4.** Let \(\sup(\cdot)\) denotes

\[
\sup(Z) := \sup_{\Theta = (\alpha^T, A^T)^T; \Theta \in \Theta} \mathbb{E}_{\alpha, \beta, F, G}[Z|X_{1}, \ldots, X_0].
\]

For every \(\Theta \in \Theta\), there exist a \(\delta > 0\) and a random variable \(M^0\) such that for all \(i, j = 1, \ldots, d, k = 1, \ldots, p\), \(\sup(|f_{\alpha}(e_i, e_j)|) \leq M^0\), \(\sup(|\hat{f}_{\alpha}(e_i, e_j)|) \leq M^0\), \(\sup(|g_{\alpha_i}(e_i, e_j)|) \leq M^0\), \(\sup(|\hat{g}_{\alpha_i}(e_i, e_j)|) \leq M^0\), and \(\mathbb{E}_{\alpha, \beta, F, G}[M^0] < \infty\).
(A5) Let $\ell_n$ and $\hat{\ell}_n$ be the first and second derivatives of log-likelihood for $\{X_p, \ldots, X_n\}$. The information equality $E_{\theta} \frac{d}{d\theta} \hat{\ell}_n |\Theta_0 | = -E_{\theta} \frac{d}{d\theta} \ell_n |\Theta_0 |$ is satisfied, and $E_{\theta} \frac{d}{d\theta} \hat{\ell}_n |\Theta_0 |$ is nonsingular and continuous in $\Theta$.

(A6) For all $\theta \in \Theta$, $\| E_{\theta} \frac{d}{d\theta} \hat{\ell}_n |\Theta_0 | \| < \infty$ and $\| E_{\theta} \frac{d}{d\theta} \hat{\ell}_n |X_0 \| < \infty$.

(A7) $F_\alpha \times G_\alpha = F_{\alpha'} \times G_{\alpha'}$ implies $(\alpha, A) = (\alpha', A')$.

To prove that $(\xi^{(n)}(u, \theta, F, G))_{n \in \mathbb{Z}}$ has the LAN property, we need to determine the behavior of a localized log-likelihood ratio. To this end, we first write down the following likelihood:

$L_n(\theta | X_p, \ldots, X_n) = \nu \{ X_p, \ldots, X_1 \} \prod_{t=0}^{n} p_{(X_{t-1}, X_t)}(X_t)$.

In addition, we introduce the following log-likelihood:

$\ell_n(\theta | X_p, \ldots, X_n) = \log L_n(\theta | X_p, \ldots, X_n)$

$= \log \nu \{ X_p, \ldots, X_1 \} + \sum_{t=0}^{n} \log p_{(X_{t-1}, X_t)}(X_t)$

$:= \log \nu \{ X_p, \ldots, X_1 \} + \sum_{t=0}^{n} \ell(\theta | X_{t-1}, \ldots, X_t)$.

Following Drost et al. (2008), we establish the LAN property by using a Taylor expansion. To do so, the transition score for $\theta$ is needed. The transition score ($\ell$) can be derived by calculating the partial derivatives of $\ell(- \log P)$ as follows. For the partial derivatives with respect to $a_i (i = 1, \ldots, d)$,

$\ell_i (x_{t-p}, \ldots, x_t; \theta) = \frac{\partial}{\partial a_i} \ell(\theta | x_{t-p}, \ldots, x_t)$

$= \frac{\partial}{\partial a_i} \log \nu \{ x_{t-p}, \ldots, x_t \} + \sum_{t=0}^{n} \frac{\partial}{\partial a_i} \log p_{(X_{t-1}, X_t)}(X_t)$

$= \frac{\partial}{\partial a_i} \log \nu \{ x_{t-p}, \ldots, x_t \} + \sum_{t=0}^{n} \ell(\theta | x_{t-1}, \ldots, x_t)$

Following Drost et al. (2008), we establish the LAN property by using a Taylor expansion. To do so, the transition score for $\theta$ is needed. The transition score ($\ell$) can be derived by calculating the partial derivatives of $\ell(- \log P)$ as follows. For the partial derivatives with respect to $a_i (i = 1, \ldots, d)$,

$\ell_i (x_{t-p}, \ldots, x_t; \theta) = \frac{\partial}{\partial a_i} \ell(\theta | x_{t-p}, \ldots, x_t)$

$= \frac{\partial}{\partial a_i} \log \nu \{ x_{t-p}, \ldots, x_t \} + \sum_{t=0}^{n} \frac{\partial}{\partial a_i} \log p_{(X_{t-1}, X_t)}(X_t)$

$= \frac{\partial}{\partial a_i} \log \nu \{ x_{t-p}, \ldots, x_t \} + \sum_{t=0}^{n} \ell(\theta | x_{t-1}, \ldots, x_t)$

Following Drost et al. (2008), we establish the LAN property by using a Taylor expansion. To do so, the transition score for $\theta$ is needed. The transition score ($\ell$) can be derived by calculating the partial derivatives of $\ell(- \log P)$ as follows. For the partial derivatives with respect to $a_i (i = 1, \ldots, d)$,

$\ell_i (x_{t-p}, \ldots, x_t; \theta) = \frac{\partial}{\partial a_i} \ell(\theta | x_{t-p}, \ldots, x_t)$

$= \frac{\partial}{\partial a_i} \log \nu \{ x_{t-p}, \ldots, x_t \} + \sum_{t=0}^{n} \frac{\partial}{\partial a_i} \log p_{(X_{t-1}, X_t)}(X_t)$

$= \frac{\partial}{\partial a_i} \log \nu \{ x_{t-p}, \ldots, x_t \} + \sum_{t=0}^{n} \ell(\theta | x_{t-1}, \ldots, x_t)$

Following Drost et al. (2008), we establish the LAN property by using a Taylor expansion. To do so, the transition score for $\theta$ is needed. The transition score ($\ell$) can be derived by calculating the partial derivatives of $\ell(- \log P)$ as follows. For the partial derivatives with respect to $a_i (i = 1, \ldots, d)$,
where $G_{i,j,k} \equiv G_{i,j,k}^{e_j,e_k}$ if $(i,j,k) \neq (i,j,k)$ and $G_{i,j,k} \equiv (x_{i+1} \cdots x_i^{e_j,e_k})$ if $(i,j,k) = (i,j,k)$.

Then, by using the Equations (3) and (4), we can derive a Taylor expansion of the localized log-likelihood ratio, and the appropriate limit theorems suggest that $(L^{(n)}(u, \theta, F, G))_{n \in \mathbb{Z}}$ has the LAN property as follows:

**Theorem 1.** Suppose that any $F_0 \times G_0 \in \mathcal{F} \times \mathcal{G}$ given any $\theta = (a^T, A^T)^T \in \Theta$ satisfies Assumptions (A1)-(A7), and let $\nu$ be a probability measure on $\mathbb{Z}^d$ with finite support. Then, the sequence of experiments $(L^{(n)}(u, \theta, F, G))_{n \in \mathbb{Z}}$ has the LAN property in $\Theta$, i.e., for every $u \in \mathbb{R}^q$ the following expansion holds,

\[
\log \frac{dP^{(n)}_{X_0, U_0, \nu(F,G)}}{dP^{(n)}_{X_0, U_0, F,G}} = \log \frac{L_n(\theta + \frac{1}{\sqrt{n}} X_{-p}, \ldots, X_n)}{L_n(\theta | X_{-p}, \ldots, X_n)} = u^T S_n - \frac{1}{2} u^T J u + R_n
\]

where the score (also called the central sequence)

\[
S_n = S_n(\theta) = \frac{1}{\sqrt{n}} \sum_{t=0}^n \left( \begin{array}{c}
\dot{\epsilon}_a(X_{t-p}, \ldots, X_t; \theta) \\
\dot{\epsilon}_A(X_{t-p}, \ldots, X_t; \theta)
\end{array} \right)_{j=1 \ldots d}
\]

satisfies

\[
S_n \overset{d}{\rightarrow} N(0, J) \quad \text{under } P_{X_0, U_0, F, G}.
\]

The Fisher information defined by

\[
J = J(\theta) = \begin{pmatrix}
(J_{a,a})_{i,j=1 \ldots d} & (J_{a,A})_{i,j=1 \ldots d, k=1 \ldots p} \\
(J_{a,A})_{i,j=1 \ldots d, k=1 \ldots p} & (J_{A,A})_{i,j=k=1 \ldots p}
\end{pmatrix}
\]

with $\dot{\epsilon} = \dot{\epsilon}(X_{-p}, \ldots, X_0; \theta)$ is nonsingular, and $R_n = R_n(u, \theta) \overset{p}{\rightarrow} 0$ under $P_{X_0, U_0, F, G}$.

### 3. Efficient estimation

This section provides efficient estimators based on the one-step update method. First, we use the multivariate conditional least squares estimator as an initial estimator of $\theta$ (e.g., Bu, McCabe, & Hadri, 2008).

**Definition 1** Suppose that $(X_{-p}, \ldots, X_0, X_1, \ldots, X_t)$ is observed from the MGINAR(p) process defined by (1). Then, the conditional least squares (CLS) estimator $\hat{\theta}_{CLS}$ for $\theta$ is defined by

\[
\hat{\theta}_{CLS} = \arg \min_{\theta_0 \in \Theta} Q_n(\theta)
\]

where

\[
Q_n(\theta) = \sum_{t=0}^T (X_t - g_t(\theta))^T (X_t - g_t(\theta)) \quad \text{and} \quad g_t(\theta) = E[X_t | F_{t-1}] = \theta + \sum_{k=1}^p A^k X_{t-k}.
\]

Note that by calculating the derivative of $Q_n$ with respect to all entries of $\theta$, we have

\[
\hat{\theta}_{CLS} = (Z^T Z)^{-1} Z^T Y.
\]
where \( Y \in \mathbb{R}^{d(n+1)} \) and \( Z \in \mathbb{R}^{d(n+1) \times d} \) are
\[
Y = (X_0^T, \ldots, X_n^T)^T \quad \text{and} \quad Z = (Z_0^T, \ldots, Z_n^T)^T \otimes I_d
\]
with \( Z_t = (1, X_0^T, \ldots, X_{t-1}^T)^T \in \mathbb{R}^{d \times 1} \).

Then, Du and Li (1991) showed the following.

**Proposition 1.**
\[
\sqrt{n} \left( \hat{\theta}_{\text{CLS}} - \theta \right) \xrightarrow{d} N(0, \Gamma^{-1} \Sigma^{-1}) \quad \text{under} \quad \mathbb{P}_{\nu, \theta, G}
\]
where
\[
\Gamma = \mathbb{E}_{\nu, \theta, G} \left[ \frac{\partial g_i(\theta)}{\partial \theta} \frac{\partial g_i(\theta)}{\partial \theta}^T \right], \quad \Sigma = \mathbb{E}_{\nu, \theta, G} \left[ (X_t - g_i(\theta))(X_t - g_i(\theta))^T \frac{\partial g_i(\theta)}{\partial \theta} \right].
\]

Moreover, we have the following.

**Proposition 2.** The CLS estimator \( \hat{\theta}_{\text{CLS}} \) is not asymptotically efficient, because it is evident that for some \( w \in \mathbb{R}^q \)
\[
w^T (\Gamma^{-1} \Sigma^{-1} - J^{-1}) w > 0 \quad \text{under} \quad \mathbb{P}_{\nu, \theta, G}
\]
except for some special cases.

Since we have a \( \sqrt{n} \)-consistent but inefficient estimator of \( \theta \), we update the CLS estimator to an efficient estimator by using the LAN result.

**Theorem 2.** Let \( \nu \) be a probability measure on \( \mathbb{Z}^d \) with finite support. Let \( \hat{\theta}_{\text{CLS}} \) be a CLS estimator. Define
\[
\hat{\theta}_* := \hat{\theta}_{\text{CLS}} + \frac{1}{\sqrt{n}} J_n(\hat{\theta}_{\text{CLS}})^{-1} S_n(\hat{\theta}_{\text{CLS}})
\]
where
\[
J_n(\theta) = \left( \begin{array}{c}
\frac{1}{n} \sum_{i=0}^n \hat{c}_i(X_i; \theta) \hat{c}_i(X_i; \theta) \\
\frac{1}{n} \sum_{i=0}^n \hat{c}_i(X_i; \theta) \hat{c}_i(X_i; \theta)
\end{array} \right)
\]
with \( \hat{c}_i(X_i; \theta) = \hat{c}_i(X_i-1, \ldots, X_1; \theta) \). Then, under Assumptions (A1)-(A7), \( \hat{\theta}_* \) is an asymptotically efficient estimator of \( \theta \) in the sequence of experiments \( (\mathcal{E}^{(n)}(\nu, \theta, F, G))_{n \in \mathbb{Z}_+} \). Moreover, \( J_n^{-1} \) yields a consistent estimator of \( J^{-1} \), i.e.,
\[
J_n(\hat{\theta}_{\text{CLS}})^{-1} \xrightarrow{P} J^{-1} \quad \text{under} \quad \mathbb{P}_{\nu, \theta, G}
\]

4. Numerical study
In this section, we first examine asymptotic relative efficiency (ARE) of our proposed estimator and the CLS estimator through some simulation experiments. Then, we present a real data analysis to illustrate the application of the proposed estimation method.

4.1. Simulation study
Our proposed estimator (\( \hat{\theta}_* \)) and the conditional least squares estimator (\( \hat{\theta}_{\text{CLS}} \)) under the MGINAR model are compared through a series of simulation experiments. Specifically, we assess the small sample properties of the two estimators in the following cases: a Binomial thinning operator and a Binomial innovation (Case 1); and a Poisson thinning operator and a Poisson innovation (Case 2) with \( d = 2 \) and \( p = 1 \). The count series \( \{X_t\} \) are defined by
Let $$\theta = (\alpha_1, \alpha_2, A_{11}, A_{12}, A_{21}, A_{22})^T = (0.09, 0.21, 0.24, 0.12, 0.18, 0.16)^T$$.

Note that the parameter vector is chosen to obtain stationary count series.

We ran 1000 Monte Carlo replicas with sample sizes $$n = 10, 100, 1000, 10000$$. For each replica, we estimate the model parameters based on two procedures (CLS: $$\hat{\theta}_{CLS}$$, Efficient Est: $$\hat{\theta}_e$$) and calculate the (approximated) bias (Table 1) and diagonal part of the (approximated) MSE (Table 2) of the parameter estimators. Finally, we calculate the (approximated) ARE (Table 3) defined by

$$\frac{\text{MSE of } \hat{\theta}_e}{\text{MSE of } \hat{\theta}_{CLS}}$$.

Simulations are carried out in R. For the calculation of the score $$\hat{\ell}$$ under the given distributions $$F$$ and $$G$$, please see the derivation of the score $$\hat{\ell}$$ for each case in the Appendix.

The bias results are reported in Table 1. It can be seen that the biases for both estimators tend to 0 when the sample size is sufficiently large. This implies that the both estimators are asymptotically unbiased. However, for the CLS estimator of $$A_{11}$$ and $$A_{22}$$, the biases are relatively large, which implies that the convergence rate is relatively slow. In contrast, our proposed estimator improves the CLS estimator in terms of bias.

The corresponding MSE results are displayed in Table 2. Similar to the bias results, the MSEs for both estimators tend to 0 when the sample size is sufficiently large. This implies that both estimators converge to the true values in probability. However, for the CLS estimator of $$A_{12}$$ and $$A_{21}$$, the MSEs are relatively large, which implies that the convergence rate of the variance is relatively slow. In contrast, our proposed estimator improves the CLS estimator, because it appears that the MSEs of all components converge to 0.

Finally, the ARE (asymptotic relative efficiency) results are given in Table 3. The ARE of two estimators is defined as the ratio of their asymptotic variances (e.g., Cox & Hinkley, 1974; Serfling, 2011). Let $$\hat{\theta}_1$$ and $$\hat{\theta}_2$$ be two estimators of $$\theta \in \Theta \subset \mathbb{R}^q$$. Let $$\Sigma_1$$ and $$\Sigma_2$$ be the asymptotic covariance matrices, i.e.,

$$\sqrt{n}(\hat{\theta}_1 - \theta) \overset{d}{=} N(0, \Sigma_1)$$
Table 1. Bias results for the MGINAR(1) model

| Parameter | True | CLS | Efficient Est |
|-----------|------|-----|---------------|
|           |      | Sample size (n) | Sample size (n) |
|           | 10   | 100 | 1000 | 10000 | 10 | 100 | 1000 | 10000 |
| Case1: Binomial thinning operator and Binomial innovation | | | | |
| $a_1$     | 0.09 | 0.0213 | 0.0008 | -0.0010 | -0.0010 | 0.0165 | -0.0020 | -0.0016 | -0.0013 |
| $a_2$     | 0.21 | 0.0469 | 0.0046 | -0.0016 | -0.0018 | 0.0530 | 0.0126 | -0.0007 | -0.0014 |
| $A_{11}$  | 0.24 | -0.0787 | -0.0228 | -0.0046 | -0.0029 | -0.0224 | -0.0188 | -0.0028 | -0.0020 |
| $A_{12}$  | 0.12 | 0.1343 | 0.0613 | 0.0593 | 0.0600 | 0.2738 | -0.0129 | -0.0058 | -0.0013 |
| $A_{21}$  | 0.18 | 0.0304 | -0.0526 | -0.0527 | -0.0532 | 0.1010 | -0.0481 | -0.0190 | -0.0103 |
| $A_{22}$  | 0.16 | 0.0261 | 0.0070 | 0.0024 | 0.0016 | 0.0643 | 0.0011 | -0.0011 | 0.0013 |
| Case2: Poisson thinning operator and Poisson innovation | | | | |
| $a_1$     | 0.09 | 0.0393 | 0.0080 | 0.0049 | 0.0042 | 0.0247 | 0.0014 | 0.0037 | 0.0037 |
| $a_2$     | 0.21 | 0.0660 | 0.0069 | 0.0042 | 0.0029 | 0.0422 | 0.0050 | 0.0042 | 0.0030 |
| $A_{11}$  | 0.24 | -0.0853 | -0.0314 | -0.0061 | 0.0006 | -0.0182 | -0.0137 | -0.0008 | 0.0040 |
| $A_{12}$  | 0.12 | 0.1463 | 0.0573 | 0.0551 | 0.0571 | 0.2552 | -0.0338 | -0.0200 | -0.0097 |
| $A_{21}$  | 0.18 | 0.0202 | -0.0493 | -0.0581 | -0.0568 | 0.0977 | -0.0507 | -0.0219 | -0.0125 |
| $A_{22}$  | 0.16 | 0.0206 | 0.0067 | -0.0014 | -0.0008 | 0.0764 | 0.0103 | -0.0013 | -0.0008 |
Table 2. Diagonal part of MSE Results for the MGINAR(1) model

| Parameter | True | CLS | Efficient Est |
|-----------|------|-----|---------------|
|           |      | Sample size \((n)\) | Sample size \((n)\) |
| \(\sigma_1\) | 0.09 | 0.0200 | 0.0013 | 0.0001 | 0.0000 | 0.0287 | 0.0013 | 0.0001 | 0.0000 |
| \(\sigma_2\) | 0.21 | 0.0432 | 0.0025 | 0.0002 | 0.0000 | 0.0553 | 0.0034 | 0.0002 | 0.0000 |
| \(A_{11}\) | 0.24 | 0.0564 | 0.0165 | 0.0020 | 0.0002 | 0.1048 | 0.0196 | 0.0019 | 0.0001 |
| \(A_{12}\) | 0.12 | 0.1293 | 0.0230 | 0.0058 | 0.0038 | 0.2460 | 0.0102 | 0.0016 | 0.0001 |
| \(A_{21}\) | 0.18 | 0.0831 | 0.0114 | 0.0038 | 0.0029 | 0.1437 | 0.0162 | 0.0019 | 0.0002 |
| \(A_{22}\) | 0.16 | 0.0317 | 0.0062 | 0.0011 | 0.0001 | 0.0677 | 0.0073 | 0.0011 | 0.0000 |

**Case 1: Binomial thinning operator and Binomial innovation**

| Parameter | True | CLS | Efficient Est |
|-----------|------|-----|---------------|
|           |      | Sample size \((n)\) | Sample size \((n)\) |
| \(\sigma_1\) | 0.09 | 0.0318 | 0.0018 | 0.0001 | 0.0000 | 0.0379 | 0.0017 | 0.0001 | 0.0000 |
| \(\sigma_2\) | 0.21 | 0.0525 | 0.0032 | 0.0003 | 0.0000 | 0.0648 | 0.0032 | 0.0003 | 0.0000 |
| \(A_{11}\) | 0.24 | 0.0613 | 0.0194 | 0.0023 | 0.0002 | 0.1302 | 0.0243 | 0.0022 | 0.0002 |
| \(A_{12}\) | 0.12 | 0.1413 | 0.0241 | 0.0056 | 0.0035 | 0.2390 | 0.0108 | 0.0026 | 0.0003 |
| \(A_{21}\) | 0.18 | 0.0885 | 0.0132 | 0.0045 | 0.0033 | 0.1447 | 0.0167 | 0.0022 | 0.0002 |
| \(A_{22}\) | 0.16 | 0.0285 | 0.0060 | 0.0011 | 0.0001 | 0.0806 | 0.0069 | 0.0011 | 0.0001 |
for $i = 1, 2$, respectively. Then, the ARE of $\hat{\theta}_2$ and $\hat{\theta}_1$ is given by

$$\text{ARE}(\hat{\theta}_2, \hat{\theta}_1) = \left( \frac{C_{18}}{C_{19}} \right)^{1/q}.$$  

In our study, we consider the ARE of our proposed estimator ($\hat{\theta}$) and the conditional least squares estimator ($\hat{\theta}_{\text{CLS}}$) as follows:

$$\text{ARE}(\hat{\theta}, \hat{\theta}_{\text{CLS}}) = \left( \frac{C_{18}}{C_{19}} \right)^{1/q}.$$  

Clearly, in this setup, if an ARE is larger than 1, it suggests that our proposed estimator improves the CLS estimator in terms of efficiency. Table 3 reports the ARE results, but $\Gamma^{-1} \Sigma^{-1} \Gamma^{-1}$ and $J^{-1}$ are replaced as the sample MSEs. It can be seen that the ARE tends to be larger than 1 as the sample size increases, which implies that our estimator improves the CLS estimator in terms of efficiency. We tried the same simulation studies under some different settings with respect to the parameter values. We omit them, but the results are similar.

4.2. Real data analysis

The data set consists of the number of cases of infectious diseases per week by prefecture for the period 2015–2018 (208 weeks) as reported by the National Institute of Infectious Diseases (NIID) in Japan (URL: https://www.niid.go.jp/niid/en/). Here we use the number of cases of “Epidemic keratoconjunctivitis (EK)” and “Aseptic meningitis (AM)” in Shimane prefecture. The sample path plot for the data in Figure 1 reveals some seasonality or periodicity, but it looks that there exist no trend and stationarity.

Figure 2 shows the sample ACF and PACF plots for each disease. The figure shows the time dependency but any long-range dependence is not observed, so it is acceptable to fit the MGINAR (1) model.

We suppose that the time series count data $\{X_t = (X_t^{(EK)}, X_t^{(AM)})^T\}$ follows

$$X_t = A \circ X_{t-1} + \epsilon_t$$

where both of the thinning operator and the innovation follows the Binomial or Poisson distribution. The CLS estimator ($\hat{a}_{\text{CLS}}, \hat{A}_{\text{CLS}}$) for the parameter $(a, A)$ is obtained as follows:

$$\begin{bmatrix} \hat{a}_{\text{CLS}} \\ \hat{A}_{\text{CLS}} \end{bmatrix} = \begin{bmatrix} 0.4050 & 0.1240 \\ 0.5534 & 0.1179 \end{bmatrix}.$$  

Next, our proposed estimators for the Binomial thinning and the Binomial innovation ($\hat{a}_{E,B}, \hat{A}_{E,B}$) and for the Poisson thinning and the Poisson innovation ($\hat{a}_{E,P}, \hat{A}_{E,P}$) are obtained as follows:

$$\begin{bmatrix} \hat{a}_{E,B} \\ \hat{A}_{E,B} \end{bmatrix} = \begin{bmatrix} 0.1813 & 0.1222 \\ 0.3404 & 0.1664 \end{bmatrix},$$

$$\begin{bmatrix} \hat{a}_{E,P} \\ \hat{A}_{E,P} \end{bmatrix} = \begin{bmatrix} 0.1703 & 0.0534 \\ 0.2575 & 0.1207 \end{bmatrix}.$$  

| Model | Sample size ($n$) | 10 | 100 | 1000 | 10000 |
|-------|------------------|----|-----|------|-------|
| Case1 | 0.6125           | 1.0055 | 1.2982 | 1.8404 |
| Case2 | 0.6253           | 1.0186 | 1.2027 | 1.5405 |
Figure 1. The number of cases of epidemic keratoconjunctivitis (left hand side) and aseptic meningitis (right hand side) per week in Shimane prefecture.

Figure 2. The sample ACF and PACF for the epidemic keratoconjunctivitis data (left hand side) and aseptic meningitis data (right hand side).
For both cases of the Binomial distribution and the Poisson distribution, $\hat{\alpha}_{CLS}$ is greatly changed by our proposed estimation. In contrast, there is not much change with respect to $\hat{\alpha}_{CLS}$.

Finally, we evaluate the goodness of fit for each estimator based on AIC. Denote AIC for an estimator $(\hat{\theta}, \hat{A})$ under the Binomial distribution by $\text{AIC}_{\text{Bin}}(\hat{\theta}, \hat{A})$, and under the Poisson distribution by $\text{AIC}_{\text{Pois}}(\hat{\alpha}, \hat{A})$. Then, we obtained

$$\text{AIC}_{\text{Bin}}(\hat{\theta}_{CLS}, \hat{A}_{CLS}) = 794.0651, \quad \text{AIC}_{\text{Bin}}(\hat{\alpha}_{EB}, \hat{A}_{EB}) = 718.821$$

$$\text{AIC}_{\text{Pois}}(\hat{\alpha}_{CLS}, \hat{A}_{CLS}) = 43240.3805, \quad \text{AIC}_{\text{Pois}}(\hat{\alpha}_{EP}, \hat{A}_{EP}) = 43081.671$$

From the above results, we can see that our proposed estimators take good performance in terms of goodness of fit.

Acknowledgements
We thank the anonymous referees for constructive comments. This work was supported by JSPS KAKENHI, Grant Number JP16K00036.

Funding
This work was supported by the JSPS [JP16K00036].

Author details
Hiroshi Shirai
E-mail: shirai@math.keio.ac.jp
Department of Mathematics, Keio University, Yokohama, Kanagawa, Japan.

Citation information
Cite this article as: Local asymptotic normality and efficient estimation for multivariate GINAR(1) process. In M. Lovric (Ed.), Encyclopedia of statistical science. Chichester: John Wiley & Sons.

References
Al-Osh, M. A., & Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. Journal of Time Series Analysis, 8(3), 261–275. doi:10.1111/j.1467-9892.1987.tb00438.x
Alzaid, A. A., & Al-Osh, M. A. (1990). An integer-valued $p$th-order autoregressive structure (INAR(p)) process. Journal of Applied Probability, 27(2), 314–324. doi:10.2307/3214650
Bu, R., McCabe, B., & Hadri, K. (2008). Maximum likelihood estimation of higher-order integer-valued autoregressive processes. Journal of Time Series Analysis, 29(6), 973–994. doi:10.1111/j.1467-9892.2008.00590.x
Cox, D. R., & Hinkley, D. (1974). Theoretical statistics. London: Chapman and Hall.
Drost, F. C., Van Den Akker, R., & Werker, B. J. M. (2008). Local asymptotic normality and efficient estimation for INAR (p) models. Journal of Time Series Analysis, 29(5), 783–801. doi:10.1111/j.1467-9892.2008.00581.x
Du, J. G., & Li, Y. (1991). The integer-valued autoregressive (INAR(p)) model. Journal of Time Series Analysis, 12(2), 129–142. doi:10.1111/j.1467-9892.1991.tb0073.x
Frank, J., & Subba Rao, T. (1993). Multivariate first-order integer-valued autoregression. Technical report No. 95, Universität Kaiserslautern.
Karlis, D., & Pedeli, X. (2013). Flexible bivariate INAR(1) processes using copulas. Communications in Statistics-Theory and Methods, 42(6), 723–740. doi:10.1080/03610926.2012.754666
Latour, A. (1997). The multivariate GINAR(p) process. Advances in Applied Probability, 29(1), 228–248. doi:10.2307/1427868
Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. Journal of Time Series Analysis, 19(4), 439–455. doi:10.1111/j.1467-9892.1998.issue-4
McKenzie, E. (1985). Some simple models for discrete variate time series. Water Resources Bulletin, 21(4), 645–650. doi:10.1111/j.1752-1688.1985.tb05379.x
McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. Advances in Applied Probability, 20(4), 822–835. doi:10.2307/1427362
Ristic, M. M., Bokouch, H. S., & Nastic, A. S. (2009). A new geometric first-order integer-valued autoregressive (GGINAR(1)) process. Journal of Statistical Planning and Inference, 139(7), 2218–2226. doi:10.1016/j.jspi.2008.10.007
Serfling, R. (2011). Asymptotic relative efficiency in estimation. In M. Lovric (Ed.), International encyclopedia of statistical science, pp. 68–82. Berlin, Heidelberg: Springer.
Weiß, C. H. (2018). An introduction to discrete-valued time series. Chichester: John Wiley & Sons.

Appendix

Proof of Theorem 1 The proof is similar to that for Theorem 1 in Drost et al. (2008).

Expansion of the log-likelihood ratio: Let $u = (u_1, \ldots, u_q)^T \in \mathbb{R}^q \setminus \{0\}$ with $q = d + d^2p$. Under Assumption (A1), we obtain by Taylor’s theorem,

$$\log \frac{L_n(\theta + \frac{u}{\sqrt{n}} | X^{-p}, \ldots, X_n)}{L_n(\theta | X^{-p}, \ldots, X_n)} = u^T S_n(\theta) - \frac{1}{2} u^T J_n(\hat{\theta}_n) u$$

where $\hat{\theta}_n$ is a random point on the line-segment between $\theta$ and $\theta + u/\sqrt{n}$ and
\[ J_n(\theta) = -\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} S_n(\theta). \] (8)

Then, we show

- Part 0: auxiliary calculations
- Part 1: \( S_n(\theta) \overset{d}{\Rightarrow} N(0, J) \) under \( \mathbb{P}_{\theta, F, G} \)
- Part 2: \( J_n(\theta) \overset{p}{\rightarrow} J \) under \( \mathbb{P}_{\theta, F, G} \)
- Part 3: non-singularity of \( J \).

In what follows, for simplicity, we write \((X_s, \ldots, X_t) := X_{st} \) and \((x_s, \ldots, x_t) := x_{st} \) for \( s, t \in \mathbb{Z} \) with \( s < t \).

**Part 0:** We first show that the existence of \( J \). To do so, we need to show for each \( \theta_i, \theta_j \in \Theta \)

\[ \mathbb{E}_{\theta, F, G} \left[ \dot{\epsilon}_{\theta_i}(X_{(t-p):t-1}) | X_{(t-p):(t-1)} \right], \mathbb{E}_{\theta, F, G} \left[ \frac{\partial}{\partial \theta} \log F_{\theta_i}(\epsilon(t) | X_{(t-p):(t-1)}) \right] = 0. \] (9)

It can be shown by Assumptions (A4) and (A6).

**Part 1:** From Equations (3) and (4), it follows that

\[ \mathbb{E}_{\theta, F, G} \left[ \dot{\epsilon}_{A_{ij}^{(k)}}(X_{(t-p):t-1}) | X_{(t-p):(t-1)} \right] = \mathbb{E}_{\theta, F, G} \left[ \frac{\partial}{\partial A_{ij}^{(k)}} \log G_{A_{ij}^{(k)}}(\epsilon_{ij}^{(k)} | X_{(t-p):(t-1)}) \right] = 0 \]

since \( \epsilon_{ij} \) and \( \epsilon_{ij}^{(k)} \) are independent of \( (X_{t-p}, \ldots, X_{t-1}) \). Let \( w \in \mathbb{R}^d \) and \( \dot{\epsilon}_{\theta} = (\dot{\epsilon}_a, \ldots, \dot{\epsilon}_a, \dot{\epsilon}_{A_{ij}^{(k)}}, \ldots, \dot{\epsilon}_{A_{ij}^{(k)}})^{T} \).

From the above equations, it follows that

\[ \mathbb{E}_{\theta, F, G} \left[ w^{T} \dot{\epsilon}_{\theta}(X_{(t-p):t-1}) | X_{(t-p):(t-1)} \right] = 0 \]

and by Part 0, it follows that

\[ \mathbb{E}_{\theta, F, G} \left[ \left\{ w^{T} \dot{\epsilon}_{\theta}(X_{(t-p):t-1}) \right\}^{2} \right] = w^{T} J w < \infty. \]

Hence, we have, by Lemma B.1 of Drost et al. (2008),

\[ \frac{1}{\sqrt{n}} \sum_{t=0}^{n} \left[ w^{T} \dot{\epsilon}_{\theta}(X_{(t-p):t-1}) \right] \overset{d}{\rightarrow} w^{T} N(0, J) \] under \( \mathbb{P}_{\theta, F, G} \).

An application of the Cramér-Wold device concludes the proof of Part 1.

**Part 2:** Assumption (A3) implies that, for fixed \( x_p, \ldots, x_0 \in \mathbb{Z}^d \) and for each \( \theta_i, \theta_j \in \Theta \), the mapping \( \Theta \rightarrow (\partial/\partial \theta) \log \dot{\epsilon}_{\theta}(X_{(t-p):t-1}) \)

is continuous, respectively. Since we have already proved (9) in Part 0, by Lemma B.3 of Drost et al. (2008), the proof is completed if we have \( J_n(\theta) \overset{p}{\rightarrow} J \) where
\[ J_o(\theta) = -\frac{1}{n} \sum_{t=0}^{n} \left( \frac{\partial}{\partial \theta_i} \ell_o(X_{(t-p):t}; \theta) - \frac{\partial}{\partial \theta_{i'}} \ell_o(X_{(t-p):t'; \theta}) \right) \]

For \( i,i' \in \{1,\ldots,d\} \) and \( t \in \{0,1,\ldots,n\} \), it follows that from Equation (3)

\[ \frac{\partial}{\partial \theta_i} \ell_o(X_{(t-p):t}; \theta) \]

\[ = \partial \theta_i \left[ \prod_{i'=1}^{d} \left( \prod_{j=1}^{d} G_{A_{ij}}^{t_{ij}} \right) \left( X_{(t-p):t} \right) \right] \frac{1}{p^0(\theta)} \left( \frac{p^0(\theta)}{p^0(\theta)} \right)^2 \]

\[ = \prod_{i'=1}^{d} \left( \prod_{j=1}^{d} G_{A_{ij}}^{t_{ij}} \right) \left( X_{(t-p):t} \right) \frac{1}{p^0(\theta)} \left( \frac{p^0(\theta)}{p^0(\theta)} \right)^2 \]

where

\[ f_o = \begin{cases} \tilde{f}_{o,i} & \text{if } i = i' \\ f_{o,i} & \text{if } i \neq i' \end{cases} \]

and

\[ \partial_{\theta_{i'}} \log F_o(\epsilon_{i'}) = \begin{cases} \frac{\partial}{\partial \theta_{i'}} \log F_o(\epsilon_{i'}) & \text{if } i = i' \\ \frac{\partial}{\partial \theta_{i'}} \log F_o(\epsilon_{i'}) - \frac{\partial}{\partial \theta_{i'}} \log F_o(\epsilon_{i'}) & \text{if } i \neq i' \end{cases} \]

which implies that

\[ \mathbb{E}_{\theta,F,G} \left[ \frac{\partial}{\partial \theta_i} \ell_o(X_{(t-p):t}; \theta) \right] = \mathbb{E}_{\theta,F,G} \left[ \frac{\partial}{\partial \theta_{i'}} \ell_o(X_{(t-p):t'; \theta}) \right] \]

where the third equation follows the independence between \( \epsilon_i \) and \( X_{(t-p):t-1} \). This result and Lemma B.3 of Droste et al. (2008) yield

\[ -\frac{1}{n} \sum_{t=0}^{n} \frac{\partial}{\partial \theta_{i'}} \ell_o(X_{(t-p):t}; \theta) \overset{p}{\rightarrow} -\mathbb{E}_{\theta,F,G} \left[ \frac{\partial}{\partial \theta_{i'}} \ell_o(X_{(t-p):t}; \theta) \right] = J_{o,i'} \]

By the same argument, we obtain
$$\frac{-1}{n} \sum_{t=0}^{n} \frac{\partial}{\partial \theta} \hat{A}_{i,j}^{(k)}(X_{(t-p)}; \theta) \rightarrow J_{A_{i,j}^{(k)}},$$

$$\frac{-1}{n} \sum_{t=0}^{n} \frac{\partial}{\partial \theta} \hat{A}_{i,j}^{(k)}(X_{(t-p)}; \theta) \rightarrow J_{A_{i,j}^{(k)}}(A_{i,j}^{(k)}).$$

This completes the proof of Part 2.

**Part 3:** The proof of the non-singularity for $J$ is provided by Assumption (A5) in the same way as Drost et al. (2008).

**Proof of Theorem 2** The proof is the same as Theorem 3.2 of Drost et al. (2008).

To calculate the one-step estimator $(\hat{\theta}_n)$, we need to know the score functions $(\hat{\ell}_{a_i})$ and $(\hat{\ell}_{A_{i,j}^{(k)}})$ for each distribution. In what follows, we show the derivation of these functions for Cases 1 and 2 in Section 4.

**Derivation of the score $\hat{\ell}$ for the Case 1** For Case 1, the probability functions $G_{A_{i,j}}$ and $F_{a_i}$ are given by

$$G_{A_{i,j}}(x) = A_{i,j}^{(1)}(1 - A_{i,j})^{1-x}, \quad F_{a_i}(x) = a_i^{(1)}(1 - a_i)^{1-x}$$

for $x \in \{0, 1\}$, which implies that the derivatives $g_{A_{i,j}}$ and $f_{a_i}$ are written as

$$g_{A_{i,j}}(x) = \left(\frac{x}{A_{i,j}} - \frac{1 - x}{1 - A_{i,j}}\right) G_{A_{i,j}}(x), \quad f_{a_i}(x) = \left(\frac{x}{a_i} - \frac{1 - x}{1 - a_i}\right) F_{a_i}(x).$$

Let

$$\text{Bin}_{(n,A)}(x) = \binom{n}{x} A^n (1 - A)^{n-x}, \quad \text{Bin}_{(n,A)}(x) = \left(\frac{x}{A} - \frac{n - x}{1 - A}\right) \text{Bin}_{(n,A)}(x)$$

for $n \in \mathbb{N}$ and $x \in \{0, \ldots, n\}$. Let $\text{Bin}_{(0,A)}(0) = 1$, $\text{Bin}_{(0,A)}(0) = 0$ and $\text{Bin}_{(0,A)}(x) = \text{Bin}_{(0,A)}(x) = 0$ for $x \neq 0$. Then, we obtain for $i,j = 1,2$,

$$p_{x_{(t-1)},x_t}^{\theta} = \sum_{(1)} \text{Bin}_{(1,A)}(n_1) \text{Bin}_{(x_{t-1},A_{i,1})}(n_2) \text{Bin}_{(x_{t-1},A_{i,2})}(n_3),$$

$$\quad \times \sum_{(2)} \text{Bin}_{(1,A)}(n_1') \text{Bin}_{(x_{t-1},A_{j,1})}(n_2') \text{Bin}_{(x_{t-1},A_{j,2})}(n_3'),$$

$$\hat{\ell}_{a_i}(x_{t-1}, x_t; \theta) = \sum_{(1)} \text{Bin}_{(1,A)}(n_1) \text{Bin}_{(x_{t-1},A_{i,1})}(n_2) \text{Bin}_{(x_{t-1},A_{i,2})}(n_3),$$

$$\quad \times \sum_{(2)} \text{Bin}_{(1,A)}(n_1') \text{Bin}_{(x_{t-1},A_{i,1})}(n_2') \text{Bin}_{(x_{t-1},A_{i,2})}(n_3') / p^{\theta}_{x_{(t-1)},x_t}.$$
The derivation of the score \( \hat{\beta} \) for Case 2 For Case 2, the probability functions \( G_{A_j} \) and \( F_{a} \) are given by

\[
G_{A_j}(x) := (A_{ij})^x \exp(-A_{ij})/x!, \quad F_{a}(x) := (a_i)^x \exp(-a_i)/x!
\]

for \( x \in \mathbb{Z}_+, \) which implies that the derivatives \( g_{A_j} \) and \( f_{a} \) are written by

\[
g_{A_j}(x) = \left( \frac{x}{A_{ij}} - 1 \right) G_{A_j}(x), \quad f_{a}(x) = \left( \frac{x}{a_i} - 1 \right) F_{a}(x).
\]

Let

\[
\text{Pois}_{(i)}(x) = \frac{x^e e^{-x}}{x!}, \quad \text{Pois}_{(i)}^{\circ}(x) = \left( \frac{x}{i} - 1 \right) \frac{x^e e^{-x}}{x!}
\]

and

\[
\text{Pois}_{(nA_j)}^{\circ}(x) = \begin{cases} 
0 & \text{if } n = 0 \\
\text{Pois}_{(i)}^{\circ}(x) & \text{if } n = 1 \\
\sum_{x_1 + x_2 = x} n \text{Pois}_{(i)}^{\circ}(x_1) \text{Pois}^{\circ}_{(n-1A_j)}(x_2) & \text{if } n > 1
\end{cases}
\]

for \( n \in \mathbb{Z}_+, A > 0, x \in \mathbb{Z}_+. \) Then, we obtain for \( i, j = 1, 2, \)

\[
p_{(x_1-1), x_2}^\theta = \sum_{(3)} \text{Pois}_{(n_1)}(n_1) \text{Pois}^{\circ}_{(x_{1-1}A_1)}(n_2) \text{Pois}^{\circ}_{(x_{2-1}A_2)}(n_3)
\]

\[
\times \sum_{(4)} \text{Pois}_{(n_1')}^{\circ}(n_1') \text{Pois}^{\circ}_{(x_{1-1}A_1)}(n_2') \text{Pois}^{\circ}_{(x_{2-1}A_2)}(n_3')
\]

\[
\hat{\beta}_{A_j}(x_{t-1}, x_t; \theta) = \sum_{(3)} \text{Pois}_{(n_1)}^{\circ}(n_1) \text{Pois}^{\circ}_{(x_{1-1}A_1)}(n_2) \text{Pois}^{\circ}_{(x_{2-1}A_2)}(n_3)
\]

\[
\times \sum_{(4)} \text{Pois}_{(n_1')}^{\circ}(n_1') \text{Pois}^{\circ}_{(x_{1-1}A_1)}(n_2') \text{Pois}^{\circ}_{(x_{2-1}A_2)}(n_3')/p_{(x_1-1), x_t}^\theta
\]

\[
\hat{\beta}_{A_j}(x_{t-1}, x_t; \theta) = \sum_{(3)} \text{Pois}_{(n_1)}(n_1) \text{Pois}_{(x_{j-1}A_j)}^{\circ}(n_{j+1}) \text{Pois}_{(x_{j+1}A_{j+1})}^{\circ}(n_{j+2})
\]

\[
\times \sum_{(4)} \text{Pois}_{(n_1')}^{\circ}(n_1') \text{Pois}^{\circ}_{(x_{j-1}A_j)}(n_{j+1}) \text{Pois}^{\circ}_{(x_{j+1}A_{j+1})}(n_{j+2})/p_{(x_{j-1}), x_t}^\theta
\]

where \( \sum_{(3)} = \sum_{n_1 + n_2 + n_3 = x_{t-1}} \) and \( \sum_{(4)} = \sum_{n_1 + n_2 + n_3 = x_{t-1}} \).
