ESAIM: PS 28 (2024) 350–365
https://doi.org/10.1051/ps/2024011

ESAIM: Probability and Statistics
www.esaim-ps.org

ERGODIC PROPERTIES OF SUBCRITICAL MULTITYPE GALTON–WATSON PROCESSES WITH IMMIGRATION

Gábor Szűcs*

Abstract. In the paper ergodic properties of multitype Galton–Watson processes are investigated in the subcritical case without further regularity assumptions. Sufficient and necessary conditions for the existence of the stationary distribution and its moments are provided. Under moment conditions geometric ergodicity and rate of converge for the moments of the process are proved. Geometric properties of the Markovian class structure are also studied.

Mathematics Subject Classification. 60J80, 60G10.

Received December 31, 2023. Accepted July 13, 2024.

1. Introduction

Galton–Watson processes are historically one of the oldest fields in the theory of stochastic processes, and they have several applications in life and computer sciences. Although the probability of extinction for the single type process was determined in the 19th century, research on the multitype version started only in the middle of the 20th century. As main references on the subject see the classical books of Athreya and Ney [1] and Mode [2].

Multitype Galton–Watson processes with immigration were introduced by Quine [3]. In the paper a necessary and sufficient condition was proved for the existence of stationary distribution in the subcritical case with positive regular mean matrix. Kaplan [4] investigated the general case under the same regularity condition on the mean matrix, and generalized the result of Quine. Our main goal in this paper is to extend these results without any assumption on the mean matrix.

Although the properties of the stationary distribution in the subcritical case were investigated in several papers, only a few results are available on the existence of its moments. Explicit formulas for the variance and the third moment were provided by Quine [3] and by Barczy et al. [5], respectively. Recently, Kevei and Wiandt [6] proved a sufficient condition for the existence for the moment of an arbitrary order. The geometric properties of the Markovian class structure is also an important topic. In several applications it is required that the stationary distribution should not be concentrated on a lower dimensional affine subspace of the state space of the process. For example, see the results by Pap and T. Szabó [7] and by Nedényi [8] on the existence of the maximum likelihood and the conditional least squares estimator for special branching processes, respectively.

In this paper we investigate the ergodic properties of the multitype Galton–Watson process with immigration. The research is limited to the subcritical case, but we do not assume any further regularity conditions. In our first theorem we study the Markovian class structure of the process. It is shown that there is a unique positive

Keywords and phrases: Galton–Watson processes, Multitype branching processes, Stationary distribution, Geometric ergodicity.

Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, 6720 Szeged, Hungary.

* Corresponding author: szucsg@math.u-szeged.hu

recurrent class which is reached with probability 1 starting from any initial distribution. Using this result we prove that a modified version of Quine's logarithmic moment condition is sufficient and necessary for the existence of a stationary distribution in case of an arbitrary subcritical Galton–Watson process. Geometric properties of the unique positive recurrent class are also investigated. In our last theorem we prove a necessary and sufficient condition for the existence of moments of the stationary distribution, and provide a rate of convergence for moments of the process.

2. Main results

Let $\mathbb{Z}_+$ stand for the set of nonnegative integers, and consider an arbitrary positive integer p. The p-type Galton–Watson process with immigration $\mathbf{X}_n = (X_{n,1}, \dots, X_{n,p})^\top$, $n \in \mathbb{Z}_+$, is a $\mathbb{Z}_+^p$-valued Markov chain defined by the recursion

$$\mathbf{X}_n = \sum_{k=1}^{X_{n-1,1}} \boldsymbol{\xi}_1(n,k) + \cdots + \sum_{k=1}^{X_{n-1,p}} \boldsymbol{\xi}_p(n,k) + \boldsymbol{\eta}(n), \qquad n = 1, 2, \dots, \tag{2.1}$$

where the $\mathbb{Z}_+^p$-valued random vectors

$$\mathbf{X}_0, \boldsymbol{\xi}_i(n,k), \boldsymbol{\eta}(n), \qquad i = 1, \dots, p, \quad n, k = 1, 2, \dots, \tag{2.2}$$

are independent of each other, the offspring variables $\boldsymbol{\xi}_i(n,k)$, $n, k = 1, 2, \dots,$ are identically distributed for every $i = 1, \dots, r$, and the innovation variables $\boldsymbol{\eta}(n)$, $n = 1, 2, \dots,$ are identically distributed. In the following we interpret the random vector $\mathbf{X}_n$ as the size of the nth generation of an underlying population having p different types of members. The offspring of any subpopulation of the process are called *1st generation offspring*, and by recursion, the *nth generation offspring* are defined as the offspring of the $(n-1)$th generation offspring. We say that a member is a *multigeneration offspring* if it is nth generation offspring with some positive integer n.

Throughout the paper we assume that the offspring variables have finite expectations, and we consider the mean matrix

$$\mathbf{M} = \big[E\boldsymbol{\xi}_1(1,1), \dots, E\boldsymbol{\xi}_p(1,1)\big]^\top.$$

It is well-know that the asymptotic properties of the Galton–Watson process depend on the spectral radius $\varrho(\mathbf{M})$ of the matrix $\mathbf{M}$. (See Mode [2] or Athreya and Ney [1] for classical results. Also, an interesting special case is provided by Foster and Ney [9] where decomposable processes are studied). The process is called subcritical, critical or supercritical if the spectral radius is smaller than 1, equal to 1 or larger than 1, respectively. In our paper we investigate only the subcritical (also known as stable) case. Note that in this case the multigeneration offspring of any member of the population die out in finitely many steps with probability 1.

For any $\mathbf{x} = (x_1, \dots, x_p)^\top \in \mathbb{R}^p$ and $\mathbf{y} = (y_1, \dots, y_p)^\top \in \mathbb{R}^p$ the notation $\mathbf{x} \le \mathbf{y}$ is understood componentwise, that is, $\mathbf{x} \le \mathbf{y}$ if and only if $x_i \le y_i$ for $i = 1, \dots, p$. The norm of the vector $\mathbf{x}$ is defined as $\|\mathbf{x}\| = |x_1| + \cdots + |x_p|$. By using the Kronecker delta symbol $\delta_{i,j}$ the system $\mathbf{e}_i = (\delta_{i,1}, \dots, \delta_{i,p})^\top$, $i = 1, \dots, p$, stands for the canonical basis of the vector space $\mathbb{R}^p$. The symbol $\mathbf{0}$ represents the vector $(0, \dots, 0)^\top \in \mathbb{R}^p$. The sum of an arbitrary matrix $\mathbf{A} = (A_{i,j})_{i,j=1,\dots,p} \in \mathbb{R}^{p \times p}$ and a real value ε is understood componentwise, that is,

$$\mathbf{A} + \varepsilon = (A_{i,j} + \varepsilon)_{i,j=1,\dots,p} \in \mathbb{R}^{p \times p}.$$

In case of an arbitrary event A the random variable $\mathbb{1}_A$ stands for the indicator of A, and for a set $B \subseteq \mathbb{R}^p$ the function $\mathbb{1}_B(\mathbf{x}) = \mathbb{1}_{\{\mathbf{x} \in B\}}$, $\mathbf{x} \in \mathbb{R}^p$, is the indicator of B. the notations $P_{\mathbf{x}}$ and $E_{\mathbf{x}}$ mean probability and expectation with respect to the condition $\{\mathbf{X}_0 = \mathbf{x}\}$.

Our first result is a statement about the class structure of the chain $\mathbf{X}_n$, $n \in \mathbb{Z}_+$.

Theorem 2.1. *If a p-type Galton–Watson process with immigration is subcritical, then it has an aperiodic communication class $\mathcal{C} \subseteq \mathbb{Z}_+^p$ such that the process reaches $\mathcal{C}$ in finitely many steps with probability 1 for any initial distribution.*

In our next theorem we provide a necessary and sufficient condition for the existence of a stationary distribution of subcritical Galton–Watson processes. Note that such a statement is known under the condition that the mean matrix $\mathbf{M}$ is positive regular, meaning that there exists a positive integer n such that all entries of $\mathbf{M}^n$ are strictly positive. In this case by the well-known result of Quine [3] a stationary distribution exists if and only if the sum $\sum_{k=1}^{\infty} \log k P(\|\boldsymbol{\eta}(1)\| = k)$ is finite. This equivalence is not valid in the case of arbitrary offspring distributions. For example, if $\mathbf{M} = \mathbf{0}$ then $\mathbf{X}_n = \boldsymbol{\eta}(n)$ for every positive n, and the distribution of the innovation variables is a stationary distribution for the process without any additional condition.

Let $M_{i,j}^{(n)}$ stands for the (i,j)th entry of the matrix $\mathbf{M}^n$, which is the expected number of type j members among the nth generation offspring of an arbitrary member of type i. Define I as the set of those types $i = 1, \dots, p$ for which there exists a type j and integers $m_0 \geq 0$ and $m \geq 1$ such that $M_{i,j}^{(m_0)} > 0$ and $M_{j,j}^{(m)} > 0$. Since the process is subcritical, the multigeneration offspring of an arbitrary member of the process die out in finitely many steps with probability 1. However, if $i \in I$, then we have $M_{i,j}^{(m_0+nm)} > 0$ for every positive integer n. This implies that a member of type i can have nth generation offspring with positive probability for any n. On the other hand, if $i \notin I$, then it can be shown by standard calculation that we have $M_{i,j}^{(p)} = 0$ for every state j. In this case the multigeneration offspring of a member of type i die out at most in p steps with probability 1.

Let $\eta_i(n)$ denote the ith component of the vector $\boldsymbol{\eta}(n)$, and consider the subpopulation of the members in the nth generation of the process which are multigeneration offspring of the innovations $\eta_i(k)$, $k = 1, 2, \dots$ It turns out that the existence of a stationary distribution requires that the size of these subpopulations converges in distribution as $n \to \infty$ for every i. If $i \in I$, then the corresponding state j provides a feedback for the consecutive generations of the offspring of i. Because of this feedback we need similar conditions on the innovation variable $\eta_i(1)$ as in the simple-type case. If $i \notin I$, then such a feedback is not present, and the multigeneration offspring of i die out at most in p steps. In this case the size of the subpopulations mentioned above is a stationary process without any additional condition on type i. This implies that the distributions of the innovation variables $\eta_i(1)$, $i \notin I$, have no effect on the existence of a stationary distribution of the Galton–Watson process.

Theorem 2.2. *Recall that I denotes the set of those types $i = 1, \dots, p$ for which there exists a type j and integers $m_0 \geq 0$ and $m \geq 1$ such that $M_{i,j}^{(m_0)} > 0$ and $M_{j,j}^{(m)} > 0$. The subcritical Galton–Watson process with immigration $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, has a stationary distribution π if and only if we have $\sum_{k=1}^{\infty} \log k P(\eta_i(1) = k) < \infty$ for every type $i \in I$.*

Note that for an arbitrary nonnegative integer valued random variable ζ the sum $\sum_{k=1}^{\infty} \log k P(\zeta = k)$ is finite if and only if the expectation $E \log(\zeta + 1)$ is finite.

Since $\mathcal{C}$ is the only closed communication class by Theorem 2.1, a subcritical Galton–Watson process has at most one positive recurrent class. This implies that the stationary distribution is unique and concentrated on $\mathcal{C}$, if it exists. From Theorem 2.1 it also follows that every subcritical Galton–Watson process is *ψ-irreducible* and *aperiodic* in the sense of Meyn and Tweedie [10]. The *maximal irreducibility measures* ψ are those probability measures on the state space $\mathbb{Z}_+^p$ which are concentrated on $\mathcal{C}$ and put positive mass at every state in this class. Furthermore, if the (unique) stationary distribution exists then the process is *positive Harris recurrent*, and Theorem 13.0.1 of [10] implies that for any $\mathbf{x} \in \mathbb{Z}_+^p$ we have

$$\sup_{B \subseteq \mathbb{Z}_+^p} \left| P_{\mathbf{x}}(\mathbf{X}_n \in B) - \pi(B) \right| \to 0, \qquad n \to \infty.$$

Under some stronger moment conditions we provide a rate for this convergence in Corollary 2.6.

Let $\widetilde{\mathbf{X}}$ stand for a random vector with distribution π. In several applications showing the linear independence of the components of $\widetilde{\mathbf{X}}$ is required. For example, assume that we want to estimate the mean matrix $\mathbf{M}$ based on some observations $\mathbf{X}_0, \dots, \mathbf{X}_n$ by using the conditional least squares method or its weighted variant. Unfortunately, these estimators may not exist for every realization of the sample. However, Nedényi [8] showed that they are well-defined with asymptotic probability 1 as $n \to \infty$ if the components of $\widetilde{\mathbf{X}}$ are linearly independent. A similar problem arose in Pap and T. Szabó [7] about the maximum likelihood estimation of the parameters of the INAR(p) process. Since the stationary distribution π puts positive mass at every state in $\mathcal{C}$, the components of $\widetilde{\mathbf{X}}$ are linearly dependent if and only if the class $\mathcal{C}$ is a subset of a lower dimensional affine subspace of $\mathbb{R}^p$. In our next theorem we provide necessary and sufficient conditions for this property.

Definition 2.3. We say that an arbitrary type i is *minor* if we have $P(\eta_j(1) = 0) = 1$ for every types j for which there exists a nonnegative integer n such that $M_{j,i}^{(n)} > 0$.

Since $M_{i,i}^{(0)} = 1$ for all types i, there is no innovation of minor types. Also, members of minor types cannot be multigeneration offspring of the innovations $\boldsymbol{\eta}(1), \boldsymbol{\eta}(2), \dots$ These imply that if i is a minor type then all members in $X_{n,i}$ are nth generation offspring of the initial population $\mathbf{X}_0$. Since the process is subcritical, all minor types vanish from the population in finitely many steps with probability one.

Theorem 2.4. *Assume that the Galton–Watson process with immigration $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, is subcritical. The communication class $\mathcal{C}$ defined in Theorem 2.1 is a subset of a lower dimensional affine subspace of $\mathbb{R}^p$ if and only if either of the following conditions holds:*

(i) Any of the types are minor.
(ii) There exists a vector $\mathbf{c} \in \mathbb{R}^p$, $\mathbf{c} \neq \mathbf{0}$, such that $\mathbf{c}^\top \boldsymbol{\xi}_i(1,1) = 0$ almost surely for every type i, and the variable $\mathbf{c}^\top \boldsymbol{\eta}(1)$ is degenerate.

If type i is minor then $\mathcal{C}$ is a subset of the linear space defined by the equation $\mathbf{e}_i^\top \mathbf{x} = 0$, $\mathbf{x} \in \mathbb{R}^p$. If (ii) holds, then $\mathcal{C}$ is a subset of the affine subspace $\mathbf{c}^\top \mathbf{x} = \mathbf{c}^\top \boldsymbol{\eta}(1)$, $\mathbf{x} \in \mathbb{R}^p$.

In our next theorem we investigate the moments of the stationary distribution π. For this goal let $\mathcal{Z}$ be the set of those states $\mathbf{x} = (x_1, \dots, x_p)^\top \in \mathbb{Z}_+^p$ for which we have $x_i = 0$ for all minor types i. Note that a minor type cannot be an offspring of any type that is not minor, and there is no innovation of minor types. This implies that $\mathcal{Z}$ is a closed subset of the state space in the sense that $P_{\mathbf{x}}(\mathbf{X}_1 \in \mathcal{Z}) = 1$ for every $\mathbf{x} \in \mathcal{Z}$. Also, by Theorem 2.4 we have $\mathcal{C} \subseteq \mathcal{Z}$, but the two sets may not coincide. For any real value $\alpha > 0$ consider the set

$$\mathcal{F}_\alpha = \big\{ f : \mathbb{Z}_+^p \to \mathbb{R} : |f(\mathbf{x})| \leq \|\mathbf{x}\|^\alpha + 1, \mathbf{x} \in \mathbb{Z}_+^p \big\}.$$

Theorem 2.5. *Assume that the subcritical Galton–Watson process $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, has a stationary distribution π, and consider a real value $\alpha > 0$. Then, the following statements are equivalent:*

(i) The distribution π has finite moment of order α, that is, $\int_{\mathbb{Z}_+^p} \|\mathbf{y}\|^\alpha \pi(d\mathbf{y}) < \infty$.
(ii) We have $E\|\boldsymbol{\eta}(1)\|^\alpha < \infty$ and $E\|\boldsymbol{\xi}_i(1,1)\|^\alpha < \infty$ for all types i that is not minor.

Furthermore, if (i) or (ii) is satisfied then there exist finite constants $a_1 > 1$ and $a_2 > 0$ such that

$$\sum_{n=0}^{\infty} a_1^n \sup_{f \in \mathcal{F}_\alpha} \left| E_{\mathbf{x}} f(\mathbf{X}_n) - \int_{\mathbb{Z}_+^p} f(\mathbf{y}) \pi(d\mathbf{y}) \right| \leq a_2 \big(\|\mathbf{x}\|^\alpha + 1 \big), \tag{2.3}$$

for every state $\mathbf{x} \in \mathcal{Z}$. Additionally, if $E\|\boldsymbol{\xi}_i(1,1)\|^\alpha < \infty$ for every type i, then (2.3) holds for every $\mathbf{x} \in \mathbb{Z}_+^p$

It is a consequence of Theorem 4 that the supremum in formula (2.3) is of rate $o(1/a_1^n)$ as $n \to \infty$, implying that $E_{\mathbf{x}} f(\mathbf{X}_n)$ converges to $\int_{\mathbb{Z}_+^p} f(\mathbf{y}) \pi(d\mathbf{y})$ for every $f \in \mathcal{F}_\alpha$ at exponential rate. Since the function $f(\mathbf{x}) = \|\mathbf{x}\|^\beta$,

$\mathbf{x} \in \mathbb{Z}_+^p$, is an element of $\mathcal{F}_\alpha$ for every $\beta \in [0, \alpha]$, we also obtain that

$$\sum_{n=0}^{\infty} a_1^n \sup_{\beta \in [0,\alpha]} \left| E_{\mathbf{x}} \|\mathbf{X}_n\|^\beta - \int_{\mathbb{Z}_+^p} \|\mathbf{y}\|^\beta \pi(d\mathbf{y}) \right| \leq a_2 \big(\|\mathbf{x}\|^\alpha + 1 \big), \quad n \to \infty.$$

This means that those moments of the process which are of order at most α converge uniformly to the related moments of the stationary distribution.

Let us note that the finiteness of the mean matrix $\mathbf{M}$ implies that $E\|\boldsymbol{\xi}_i(1,1)\|^\alpha < \infty$ holds for every type i and for every $\alpha \in (0, 1]$. Also, the indicator function $\mathbb{1}_B$ of an arbitrary set $B \subseteq \mathbb{Z}_+^p$ is an element of $\mathcal{F}_\alpha$, and we have $E_{\mathbf{x}} \mathbb{1}_B(\mathbf{X}_n) = P_{\mathbf{x}}(\mathbf{X}_n \in B)$ and $\int_{\mathbb{Z}_+^p} \mathbb{1}_B(\mathbf{y}) \pi(d\mathbf{y}) = \pi(B)$. These facts along with Theorem 2.5 immediately implies the following statement.

Corollary 2.6. *Assume that the Galton–Watson process with immigration $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, is subcritical and $E\|\boldsymbol{\eta}(1)\|^\alpha < \infty$ with some $\alpha > 0$. Then,*

$$\sum_{n=0}^{\infty} a_1^n \sup_{B \subseteq \mathbb{Z}_+^p} \big| P_{\mathbf{x}}(\mathbf{X}_n \in B) - \pi(B) \big| < \infty, \qquad \mathbf{x} \in \mathbb{Z}_+^p,$$

meaning that the process is geometrically ergodic.

As a final remark we note that inequality (2.3) can be stated in an unconditional form too, where the initial value of the process is not fixed. If condition (ii) of Theorem 2.5 is satisfied, the distribution of $\mathbf{X}_0$ is concentrated on the set $\mathcal{Z}$, and $E\|\mathbf{X}_0\|^\alpha < \infty$, then by conditioning with respect to $\mathbf{X}_0$ we obtain the inequality

$$\sum_{n=0}^{\infty} a_1^n \sup_{f \in \mathcal{F}_\alpha} \left| Ef(\mathbf{X}_n) - \int_{\mathbb{Z}_+^p} f(\mathbf{y}) \pi(d\mathbf{y}) \right| \leq a_2 \big(E\|\mathbf{X}_0\|^\alpha + 1 \big) < \infty.$$

Furthermore, if all of the offspring variables have finite moment of order α, then the restriction of the initial distribution to $\mathcal{Z}$ can be omitted. Using this result one can show the uniform convergence of the unconditional probabilities $P(\mathbf{X}_n \in B)$, $B \subseteq \mathbb{Z}_+^p$, similarly as of the conditional ones in Corollary 2.6.

3. Proofs

In this section, we present the proofs of the results stated in Section 2.

Proof of Theorem 2.1. We will use the representation of the multitype Galton–Watson process $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, provided in Section 2.7 of Mode [2]. Let the vectors $\mathbf{Y}_n$ and $\mathbf{V}_{k+n}(k)$, $n, k = 1, 2, \ldots$, stand for the number of the nth generation offspring of the initial population $\mathbf{X}_0$ and of the innovation variable $\boldsymbol{\eta}(k)$, respectively. Also, let $\mathbf{Y}_0 = \mathbf{X}_0$ and $\mathbf{V}_n(n) = \boldsymbol{\eta}(n)$ for every n. Then, we obtain the representation of [2] in the form

$$\mathbf{X}_n = \mathbf{Y}_n + \mathbf{Z}_n = \mathbf{Y}_n + \mathbf{V}_n(1) + \cdots + \mathbf{V}_n(n), \qquad n = 1, 2, \ldots, \tag{3.1}$$

and the independence of the variables in (2.2) implies that $\mathbf{Y}_n, \mathbf{V}_n(1), \ldots, \mathbf{V}_n(n)$ are independent of each other. (This equation can be proved by standard calculations too, by showing that the probability generation function of $\mathbf{X}_n$ is equal to the product of the probability generating functions of the variables on the right side). From the definitions of the variables it follows that the sequences $\mathbf{Z}_n = \mathbf{V}_n(1) + \cdots + \mathbf{V}_n(n)$, $n = 1, 2, \ldots$, and $\mathbf{Y}_n$, $n \in \mathbb{Z}_+$, are independent of each other. Let us note that $\mathbf{Y}_n$, $n \in \mathbb{Z}_+$, is a multitype Galton–Watson process without immigration, which implies that this process becomes extinct in finitely many steps with probability 1 starting from any initial distribution.

For $n = 1, 2, \dots$ let $\mathcal{D}_n \subseteq \mathbb{Z}_+^p$ denote the range of the variable $\mathbf{Z}_n$, that is, the set of those states $\mathbf{x} \in \mathbb{Z}_+^p$ for which $P(\mathbf{Z}_n = \mathbf{x}) > 0$. Since $\mathbf{Z}_n$ and $\mathbf{Z}_{n+1}$ are independent of $\mathbf{X}_0$, and on the event $\{\mathbf{X}_0 = \mathbf{0}\}$ we have $\mathbf{X}_n = \mathbf{Z}_n$ and $\mathbf{X}_{n+1} = \mathbf{Z}_{n+1}$, we get that

$$\begin{aligned} 0 &= P\big(\mathbf{Z}_{n+1} \notin \mathcal{D}_{n+1} \mid \mathbf{X}_0 = \mathbf{0}\big) = P\big(\mathbf{X}_{n+1} \notin \mathcal{D}_{n+1} \mid \mathbf{X}_0 = \mathbf{0}\big) \\ &= \sum_{\mathbf{x} \in \mathbb{Z}_+^p} P\big(\mathbf{X}_{n+1} \notin \mathcal{D}_{n+1} \mid \mathbf{X}_n = \mathbf{x}, \mathbf{X}_0 = \mathbf{0}\big) P\big(\mathbf{X}_n = \mathbf{x} \mid \mathbf{X}_0 = \mathbf{0}\big) \\ &= \sum_{\mathbf{x} \in \mathcal{D}_n} P_{\mathbf{x}}(\mathbf{X}_1 \notin \mathcal{D}_{n+1}) P(\mathbf{Z}_n = \mathbf{x}). \end{aligned}$$

Because the terms of the last sum are nonnegative, it follows from the definition of $\mathcal{D}_n$ that $P_{\mathbf{x}}(\mathbf{X}_1 \notin \mathcal{D}_{n+1}) = 0$ for every $\mathbf{x} \in \mathcal{D}_n$. It is a consequence that the set $\mathcal{C}_n := \cup_{k=n}^{\infty} \mathcal{D}_k \subseteq \mathbb{Z}_+^p$ is closed for any n in the sense that $P_{\mathbf{x}}(\mathbf{X}_1 \in \mathcal{C}_n) = 1$ holds for every $\mathbf{x} \in \mathcal{C}_n$.

Let us recall that the sequence $\mathbf{Y}_n$, $n \in \mathbb{Z}_+$, is a subcritical Galton–Watson process without immigration, which implies that $P_{\mathbf{x}}(\mathbf{Y}_n = \mathbf{0}) \to 1$ as $n \to \infty$ for any $\mathbf{x} \in \mathbb{Z}_+^p$. Let $n^*(\mathbf{x})$ stand for the smallest integer such that $P_{\mathbf{x}}(\mathbf{Y}_n = \mathbf{0}) > 0$ holds for every $n \geq n^*(\mathbf{x})$. Because the number of the multigeneration offspring of the members of the initial population are independent of each other, we obtain that

$$P_{\mathbf{x}}(\mathbf{Y}_n = \mathbf{0}) = P_{\mathbf{e}_1}(\mathbf{Y}_n = \mathbf{0})^{x_1} \cdots P_{\mathbf{e}_p}(\mathbf{Y}_n = \mathbf{0})^{x_n} > 0$$

for every $n \geq n^*$, where $n^* = \max(n^*(\mathbf{e}_1), \dots, n^*(\mathbf{e}_p))$. That is, the process $\mathbf{Y}_n$, $n \in \mathbb{Z}_+$, dies out in n^* steps with positive probability starting from any initial state $\mathbf{x}$.

Let $\mathcal{C} = \mathcal{C}_{n^*}$, and consider an arbitrary integer $n \geq n^*$ and states $\mathbf{x} \in \mathbb{Z}_+^p$, $\mathbf{z} \in \mathcal{D}_n$. Since $\mathbf{Z}_n$ is independent of the variables $\mathbf{X}_0 = \mathbf{Y}_0$ and $\mathbf{Y}_n$, we obtain that

$$P_{\mathbf{x}}(\mathbf{X}_n = \mathbf{z}) \geq P_{\mathbf{x}}\big(\mathbf{Y}_n = \mathbf{0}, \mathbf{Z}_n = \mathbf{z}\big) = P_{\mathbf{x}}(\mathbf{Y}_n = \mathbf{0}) P(\mathbf{Z}_n = \mathbf{z}) > 0, \tag{3.2}$$

meaning that every elements of $\mathcal{D}_n$ are accessible from the arbitrary state $\mathbf{x}$ in n steps. As a consequence, the elements of $\mathcal{C}$ communicate with each other. Since $\mathcal{C}$ is closed, it is a communication class of the process $\mathbf{X}_n$, $n \in \mathbb{Z}_+$. Consider an arbitrary state $\mathbf{z} \in \mathcal{D}_n$ and a nonnegative integer m, and let $\mathbf{x} \in \mathbb{Z}_+^p$ be a state such that $P_{\mathbf{z}}(\mathbf{X}_m = \mathbf{x}) > 0$. Using equation (3.2) again we find that

$$P_{\mathbf{z}}(\mathbf{X}_{n+m} = \mathbf{z}) \geq P_{\mathbf{z}}(\mathbf{X}_m = \mathbf{x}) P_{\mathbf{x}}(\mathbf{X}_n = \mathbf{z}) > 0,$$

that is, state $\mathbf{z}$ is accessible from itself in $n + m$ steps. Since m was arbitrary nonnegative integer, the communication class $\mathcal{C}$ is aperiodic.

To prove the theorem it is only remains to show that the process $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, reaches the class $\mathcal{C}$ in finitely many steps with probability 1 starting from any initial distribution. Because the state space is countable, it is enough to prove this statement under the condition $\{\mathbf{X}_0 = \mathbf{x}\}$, where $\mathbf{x} \in \mathbb{Z}_+^p$ is an arbitrary fixed state. Note that $\{\mathbf{Y}_n = \mathbf{0}\}$, $n \in \mathbb{Z}_+$, is an increasing sequence of events. From this we get that

$$\begin{aligned} P_{\mathbf{x}}\big(\exists n \geq n^* : \mathbf{X}_n \in \mathcal{C}\big) &\geq P_{\mathbf{x}}\big(\exists n \geq n^* : \mathbf{X}_n \in \mathcal{D}_n\big) \geq P_{\mathbf{x}}\big(\exists n \geq n^* : \mathbf{Y}_n = \mathbf{0}\big) \\ &= P_{\mathbf{x}}\big(\cup_{n=n^*}^{\infty} \{\mathbf{Y}_n = \mathbf{0}\}\big) = \lim_{n \to \infty} P_{\mathbf{x}}(\mathbf{Y}_n = \mathbf{0}) = 1, \end{aligned}$$

which completes the proof. □

In the next step we prove Theorem 2.2. For this goal we need some technical results stated in Propositions 3.1–3.3.

Proposition 3.1. *Consider independent and identically distributed nonnegative valued random variables* $\xi_1, \xi_2 \dots$ *such that* $E\xi_1 \in (0,\infty)$. *Let* η *be a nonnegative integer valued random variable being independent of* $\xi_1, \xi_2 \dots$. *Then,*

$$E\log\left(\sum_{k=1}^{\eta}\xi_k+1\right)<\infty \qquad \text{if and only if} \qquad E\log(\eta+1)<\infty.$$

Proof. Assume that $E\log(\eta+1)$ is finite. By conditioning with respect to η and using Jensen's inequality to the logarithm function we get that

$$\begin{aligned}E\log\left(\sum_{k=1}^{\eta}\xi_k+1\right)&=EE\left[\log\left(\sum_{k=1}^{\eta}\xi_k+1\right)\Big|\,\eta\right]\le E\log\left(E\left[\sum_{k=1}^{\eta}\xi_k\,\Big|\,\eta\right]+1\right)\\&=E\log\big(\eta E\xi_1+1\big)\le\begin{cases}E\log(\eta+1)<\infty, & \text{if } E\xi_1\le 1,\\ \log E\xi_1+E\log(\eta+1)<\infty, & \text{if } E\xi_1\ge 1.\end{cases}\end{aligned}$$

For the contrary direction, consider the case when $E\log(\eta+1)$ is infinite. By the assumptions there exists a constant $c\in(0,1)$ such that $p=P(\xi_1\ge c)>0$. Since we have $\xi_k\ge c\mathbb{1}_{\{\xi_k\ge c\}}$ with probability 1 for every k, we obtain the inequalities

$$E\log\left(\sum_{k=1}^{\eta}\xi_k+1\right)\ge E\log\left(\sum_{k=1}^{\eta}c\mathbb{1}_{\{\xi_k\ge c\}}+1\right)\ge\log c+E\log\left(\sum_{k=1}^{\eta}\mathbb{1}_{\{\xi_k\ge c\}}+1\right). \tag{3.3}$$

Let ζ_n stand for a random variable having binomial distribution with parameters n and p. Chebishev's inequality implies that

$$P\big(\zeta_n\ge np-n^{1/2}\big)\ge P\big(|\zeta_n-E\zeta_n|\le n^{1/2}\big)\ge 1-\frac{np(1-p)}{n}\ge\frac{3}{4}.$$

Because the conditional distribution of the sum $\sum_{k=1}^{\eta}\mathbb{1}_{\{\xi_k\ge c\}}$ with respect to the event $\{\eta=n\}$ is the same as the law of ζ_n, we get that

$$E\log\left(\sum_{k=1}^{\eta}\mathbb{1}_{\{\xi_k\ge c\}}+1\right)=\sum_{n=0}^{\infty}E\log(\zeta_n+1)P(\eta=n)\ge\frac{3}{4}\sum_{n=0}^{\infty}\log\big(np-n^{1/2}+1\big)P(\eta=n). \tag{3.4}$$

If n is large enough, then we have

$$\log\big(np-n^{1/2}+1\big)\ge\log\big((n+1)p/2\big)=\log(n+1)+\log(p/2).$$

Since the expectation $E\log(\eta+1)$ is infinite, the sum on the right side of (3.4) is divergent. Hence, the proposition is proved by inequality (3.3). □

Proposition 3.2. *Let* $\mathbf{X}_n$ *and* $\mathbf{X}'_n$, $n\in\mathbb{Z}_+$, *be irreducible time-homogeneous Markov chains on some state spaces* $\mathcal{C}\subseteq\mathbb{Z}_+^p$ *and* $\mathcal{C}'\subseteq\mathbb{Z}_+^p$. *Assume that there exist states* $\mathbf{x}_0\in\mathcal{C}$, $\mathbf{x}'_0\in\mathcal{C}'$, $\mathbf{x}_0\le\mathbf{x}'_0$, *such that the variables* $\mathbf{X}_1,\mathbf{X}_2,\dots$ *are conditionally independent of* $\mathbf{X}'_0$ *with respect to the event* $\{\mathbf{X}_0=\mathbf{x}_0\}$, *and* $\mathbf{X}'_1,\mathbf{X}'_2,\dots$ *are conditionally independent of* $\mathbf{X}_0$ *with respect to* $\{\mathbf{X}'_0=\mathbf{x}'_0\}$. *Furthermore, assume that*

$$P\big(\mathbf{X}_n\le\mathbf{X}'_n\mid\mathbf{X}_0=\mathbf{x}_0,\mathbf{X}'_0=\mathbf{x}'_0\big)=1,\qquad n=1,2,\dots \tag{3.5}$$

Then the following holds:

(i) If $\mathbf{X}'_n$, $n \in \mathbb{Z}_+$, is recurrent then $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, is recurrent.
(ii) If $\mathbf{X}'_n$, $n \in \mathbb{Z}_+$, is positive recurrent then $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, is positive recurrent.

Proof. Let $p^{(n)}_{\mathbf{X}}(\cdot,\cdot)$ and $p^{(n)}_{\mathbf{X}'}(\cdot,\cdot)$ denote the n-step transition probabilities of the processes, and let P_A stand for the conditional probability with respect to $A = \{\mathbf{X}_0 = \mathbf{x}_0, \mathbf{X}'_0 = \mathbf{x}'_0\}$. Also, introduce the finite set $\mathcal{C}_0 = \{\mathbf{x} \in \mathcal{C} : \mathbf{x} \leq \mathbf{x}'_0\}$. In our proof we will use the well-known characterization of the types of states of Markov chains based on the asymptotic behavior of the transition probabilities. (See the main theorem in Section XV.5 of Feller [11], for example).

If the chain $\mathbf{X}'_n$, $n \in \mathbb{Z}_+$, is recurrent then the assumptions and the characterization of recurrent states imply that

$$\sum_{\mathbf{x}\in\mathcal{C}_0}\sum_{n=0}^{\infty} p^{(n)}_{\mathbf{X}}(\mathbf{x}_0,\mathbf{x}) = \sum_{n=0}^{\infty} P_A(\mathbf{X}_n \in \mathcal{C}_0) \geq \sum_{n=0}^{\infty} P_A(\mathbf{X}'_n = \mathbf{x}'_0) = \sum_{n=0}^{\infty} p^{(n)}_{\mathbf{X}'}(\mathbf{x}'_0,\mathbf{x}'_0) = \infty.$$

Hence, there exists a state $\mathbf{x}^* \in \mathcal{C}_0$ such that $\sum_{n=0}^{\infty} p^{(n)}_{\mathbf{X}}(\mathbf{x}_0,\mathbf{x}^*) = \infty$. Since the process $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, is irreducible, we have $p^{(k)}_{\mathbf{X}}(\mathbf{x}^*,\mathbf{x}_0) > 0$ for some $k \in \mathbb{Z}_+$. This leads to the inequality

$$\sum_{n=0}^{\infty} p^{(n+k)}_{\mathbf{X}}(\mathbf{x}_0,\mathbf{x}_0) \geq \sum_{n=0}^{\infty} p^{(n)}_{\mathbf{X}}(\mathbf{x}_0,\mathbf{x}^*) p^{(k)}_{\mathbf{X}}(\mathbf{x}^*,\mathbf{x}_0) = \infty,$$

meaning that $\mathbf{x}_0$ is a recurrent state of the chain $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, and the first statement is proved.

Similarly, if the process $\mathbf{X}'_n$, $n \in \mathbb{Z}_+$, is positive recurrent, then

$$\sum_{\mathbf{x}\in\mathcal{C}_0} \limsup_{n\to\infty} p^{(n)}_{\mathbf{X}}(\mathbf{x}_0,\mathbf{x}) \geq \limsup_{n\to\infty} P_A(\mathbf{X}_n \in \mathcal{C}_0) \geq \limsup_{n\to\infty} P_A(\mathbf{X}'_n = \mathbf{x}'_0)$$
$$= \limsup_{n\to\infty} p^{(n)}_{\mathbf{X}'}(\mathbf{x}'_0,\mathbf{x}'_0) > 0.$$

This implies that $\limsup_{n\to\infty} p^{(n)}_{\mathbf{X}}(\mathbf{x}_0,\mathbf{x}^*) > 0$ for some state $\mathbf{x}^* \in \mathcal{C}_0$. Again, if $k \in \mathbb{Z}_+$ is a constant such that $p^{(k)}_{\mathbf{X}}(\mathbf{x}^*,\mathbf{x}_0) > 0$ then

$$\limsup_{n\to\infty} p^{(n+k)}_{\mathbf{X}}(\mathbf{x}_0,\mathbf{x}_0) \geq \limsup_{n\to\infty} p^{(n)}_{\mathbf{X}}(\mathbf{x}_0,\mathbf{x}^*) p^{(k)}_{\mathbf{X}}(\mathbf{x}^*,\mathbf{x}_0) > 0.$$

From this inequality the characterization of the types implies that the chain $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, is positive recurrent, and the proof is complete. □

Proposition 3.3. *Consider subcritical p-type Galton–Watson processes $\mathbf{X}_0, \mathbf{X}_1, \ldots$ and $\mathbf{X}'_0, \mathbf{X}'_1, \ldots$ based on the offspring and innovation vectors $\boldsymbol{\xi}_i(n,k)$, $\boldsymbol{\eta}(n)$ and $\boldsymbol{\xi}'_i(n,k)$, $\boldsymbol{\eta}'(n)$, $i = 1,\ldots,p$, $n,k = 1,2,\ldots$, respectively. Assume that all of these offspring and innovation vectors are independent of $\mathbf{X}_0$ and $\mathbf{X}'_0$, and assume that $\boldsymbol{\xi}_i(n,k) \leq \boldsymbol{\xi}'_i(n,k)$ and $\boldsymbol{\eta}(n) \leq \boldsymbol{\eta}'(n)$ hold for all possible i, n and k with probability 1. Under these assumptions if the process $\mathbf{X}'_0, \mathbf{X}'_1, \ldots$ has a stationary distribution then $\mathbf{X}_0, \mathbf{X}_1, \ldots$ has a stationary distribution, as well.*

Proof. Let $\mathcal{C}$ and $\mathcal{C}'$ stand for the unique closed communication classes of the processes provided by Theorem 2.1, and consider arbitrary values $\mathbf{x}_0, \mathbf{x}'_0 \in \mathbb{Z}^p_+$ such that $\mathbf{x}_0 \leq \mathbf{x}'_0$. By the branching mechanism the processes satisfy the independence conditions of Proposition 3.2, and it can be shown by recursion with respect to n that (3.5) also holds. Since the process $(\mathbf{X}_n, \mathbf{X}'_n)$, $n \in \mathbb{Z}_+$, reaches the set $\mathcal{C} \times \mathcal{C}'$ in finitely many steps with probability 1, formula (3.5) implies that there exists states $\mathbf{x} \in \mathcal{C}$ and $\mathbf{x}' \in \mathcal{C}'$ such that $\mathbf{x} \leq \mathbf{x}'$. This means that the initial values $\mathbf{x}_0$ and $\mathbf{x}'_0$ can be chosen as elements of these classes, respectively. Since $\mathcal{C}$ and $\mathcal{C}'$ are closed, we can restrict the processes to these sets, resulting that the restricted processes satisfy all conditions of Proposition 3.2.

If $\mathbf{X}'_0, \mathbf{X}'_1, \ldots$ has a stationary distribution then this distribution must be concentrated to $\mathcal{C}'$, the only closed communication class. This implies that the restriction of the process to $\mathcal{C}'$ is a positive recurrent Markov chain. By Proposition 3.2 the process $\mathbf{X}_0, \mathbf{X}_1, \ldots$ is positive recurrent on $\mathcal{C}$, and by the theory of Markov chains the latter process has a stationary distribution. □

Remark 3.4. Consider a single-type Galton–Watson process $X'_0, X'_1, \ldots$ defined by

$$X'_n = \sum_{k=1}^{X'_{n-1}} \xi'(n,k) + \eta'(n), \qquad n = 1, 2, \ldots \tag{3.6}$$

such that $E\xi'(1,1) > 0$ and $E\log(\eta'(1)+1) = \infty$. Then, it can be shown by using classical results on Galton–Watson processes that $X'_0, X'_1, \ldots$ does not have any stationary distribution. For example, this result was proved in the subcritical case by Foster and Williamson [12] in Corollary 2. To illustrate the application of our Proposition 3.3 we present a short and simple proof for the critical and supercritical cases.

Let $X_0, X_1, \ldots$ denote the single-type Galton–Watson process corresponding to the initial value $X_0 = X'_0$ and to the offspring and innovation variables $\xi(n,k) = \mathbb{1}_{\{\xi'(n,k)\geq 1\}}$ and $\eta(n) = \eta'(n)$, $n, k = 1, 2, \ldots$ This process is defined by replacing the vectors $\xi'(n,k)$ and $\eta'(n)$ in the recursion (3.6) by $\xi(n,k)$ and $\eta(n)$, respectively. Note that the processes X_n and X'_n, $n \in \mathbb{Z}_+$, satisfy the assumptions of Proposition 3.3, and the moment conditions imply that $E\xi(1,1) > 0$ and $E\log(\eta(1)+1) = \infty$. If $\xi(1,1) = 0$ with positive probability then $E\xi(1,1) < 1$, meaning that $X_0, X_1, \ldots$ is subcritical. From this the referred result of Foster and Williamson [12] implies that $X_0, X_1, \ldots$ does not have any stationary distribution. If $\xi(1,1) = 1$ with probability 1 then $X_n \to \infty$ almost surely as $n \to \infty$, resulting that $X_0, X_1, \ldots$ does not have any stationary distribution in this case neither. Then, by using Proposition 3.3 we immediately obtain that the original process $X'_0, X'_1, \ldots$ does not have any stationary distribution.

Proof of Theorem 2.2. First, we show the existence of a stationary distribution under the logarithmic moment condition of the theorem in the case when the set I contains all types. For this goal we define an auxiliary process $\mathbf{X}'_n$, $n \in \mathbb{Z}_+$, with positive regular mean matrix for which the existence of a stationary distribution follows from standard results. Let us recall that the eigenvalues are continuous functions of the matrix entries. Since the process $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, is subcritical, there exists an $\varepsilon > 0$ such that $\varrho(\mathbf{M}') < 1$ with the positive matrix $\mathbf{M}' = \mathbf{M} + \varepsilon$. Let $\mathbf{1} \in \mathbb{R}^p$ denote the vector whose components are equal to 1, and consider random vectors $\mathbb{1}_i(n,k)$, $i = 1, \ldots, p$, $n, k = 1, 2, \ldots$, being independent of each other and of the variables in formula (2.2) and having common distribution $P(\mathbb{1}_i(n,k) = \mathbf{1}) = \varepsilon$ and $P(\mathbb{1}_i(n,k) = \mathbf{0}) = 1 - \varepsilon$. Additionally, consider the multitype Galton–Watson process $\mathbf{X}'_n$, $n \in \mathbb{Z}_+$, defined by replacing the variables (2.2) in the recursion (2.1) by the initial value $\mathbf{X}'_0 = \mathbf{X}_0$ and by the offspring and innovation variables

$$\boldsymbol{\xi}'_i(n,k) = \boldsymbol{\xi}_i(n,k) + \mathbb{1}_i(n,k), \qquad \boldsymbol{\eta}'(n) = \boldsymbol{\eta}(n), \qquad i = 1, \ldots, p, \quad n, k = 1, 2, \ldots$$

Note that the mean matrix of the new process is $E[\boldsymbol{\xi}'_1(1,1), \ldots, \boldsymbol{\xi}'_p(1,1)]^\top = \mathbf{M}'$. For any $a_1, \ldots, a_p \geq 0$ we have the algebraic inequality

$$\log(a_1+1) + \cdots + \log(a_n+1) = \log\big((a_1+1)\cdots(a_n+1)\big) \geq \log(a_1 + \cdots + a_n + 1).$$

From this we obtain that

$$E\log\big(\|\boldsymbol{\eta}(1)\| + 1\big) \leq \sum_{i=1}^{p} E\log\big(\eta_i(1) + 1\big) < \infty, \tag{3.7}$$

meaning that $\sum_{k=1}^{\infty} \log k P(\|\boldsymbol{\eta}(1)\| = k)$ is finite. Then, Corollary 1 of Kaplan [4] implies that the process $\mathbf{X}'_n$, $n \in \mathbb{Z}_+$, has a stationary distribution. Since the offspring and the innovation variables satisfy the conditions of our Proposition 3.3, it follows that the process $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, has a stationary distribution too.

Now, assume that the logarithmic moment condition of the theorem holds, and consider the case when I does not contain all types. Let I^c stand for the complement of set I, and define the random vectors $\boldsymbol{\eta}^I(n)$ and $\boldsymbol{\eta}^{I^c}(n)$, $n = 1, 2, \dots$, by their ith components

$$\eta_i^I(n) = \begin{cases} \eta_i(n), & i \in I, \\ 0, & i \in I^c, \end{cases} \qquad \eta_i^{I^c}(n) = \begin{cases} 0, & i \in I, \\ \eta_i(n), & i \in I^c, \end{cases} \qquad i = 1, \dots, p.$$

Let $\mathbf{V}_{k+n}^I(k)$ and $\mathbf{V}_{k+n}^{I^c}(k)$, $n = 0, 1, \dots$, denote the number of the nth generation offspring of the innovation variables $\boldsymbol{\eta}^I(k)$ and $\boldsymbol{\eta}^{I^c}(k)$, respectively, and introduce the $\mathbb{Z}_+^p$-valued processes

$$\mathbf{Z}_n^I = \mathbf{V}_n^I(1) + \cdots + \mathbf{V}_n^I(n), \qquad \mathbf{Z}_n^{I^c} = \mathbf{V}_n^{I^c}(1) + \cdots + \mathbf{V}_n^{I^c}(n), \qquad n = 1, 2, \dots \tag{3.8}$$

Based on the construction, the sequences in (3.8) are multitype Galton–Watson processes corresponding to the initial states $\mathbf{Z}_0^I = \mathbf{Z}_0^{I^c} = \mathbf{0}$ and to the innovation variables $\boldsymbol{\eta}^I(1), \boldsymbol{\eta}^I(2), \dots$ and $\boldsymbol{\eta}^{I^c}(1), \boldsymbol{\eta}^{I^c}(2), \dots$, respectively, having the same offspring distributions as the original process $\mathbf{X}_n$, $n \in \mathbb{Z}_+$. Also, we have $\mathbf{V}_n^I(k) + \mathbf{V}_n^{I^c}(k) = \mathbf{V}_n(k)$ for every n and k, implying the identity $\mathbf{Z}_n = \mathbf{Z}_n^I + \mathbf{Z}_n^{I^c}$, $n \in \mathbb{Z}_+$.

Consider random pairs $(\mathbf{U}_n^I, \mathbf{U}_n^{I^c})$, $n = 1, 2, \dots$, which are independent of each other and of the initial variable $\mathbf{X}_0$ such that $(\mathbf{U}_n^I, \mathbf{U}_n^{I^c})$ has the same distribution as $(\mathbf{V}_n^I(1), \mathbf{V}_n^{I^c}(1))$ for every n. Note that the pairs $(\mathbf{V}_n^I(k), \mathbf{V}_n^{I^c}(k))$, $k = 1, \dots, n$, are independent of each other, and $(\mathbf{V}_n^I(k), \mathbf{V}_n^{I^c}(k))$ has the same distribution as $(\mathbf{V}_{n-k+1}^I(1), \mathbf{V}_{n-k+1}^{I^c}(1))$. Then, for every n the joint distribution of $(\mathbf{V}_n^I(k), \mathbf{V}_n^{I^c}(k))$, $k = 1, \dots, n$, is the same as the joint distribution of $(\mathbf{U}_{n-k+1}^I, \mathbf{U}_{n-k+1}^{I^c})$, $k = 1, \dots, n$. This implies that

$$\begin{bmatrix} \mathbf{Z}_n^I \\ \mathbf{Z}_n^{I^c} \end{bmatrix} = \sum_{k=1}^{n} \begin{bmatrix} \mathbf{V}_n^I(k) \\ \mathbf{V}_n^{I^c}(k) \end{bmatrix} \stackrel{\mathcal{D}}{=} \sum_{k=1}^{n} \begin{bmatrix} \mathbf{U}_k^I \\ \mathbf{U}_k^{I^c} \end{bmatrix} \to \sum_{k=1}^{\infty} \begin{bmatrix} \mathbf{U}_k^I \\ \mathbf{U}_k^{I^c} \end{bmatrix}, \qquad n \to \infty, \tag{3.9}$$

where the convergence is understood in almost sure sense.

Since all components of the innovation variable $\boldsymbol{\eta}^I(1)$ have finite logarithmic moment by assumption, the first step of the current proof implies that the Galton–Watson process $\mathbf{Z}_n^I$, $n \in \mathbb{Z}_+$, has a (proper) stationary distribution. Additionally, the process converges to $\sum_{n=1}^{\infty} \mathbf{U}_n^I$ in distribtution by formula (3.9). From these we obtain that the law of the limit is the stationary distribution of the process, resulting that $\sum_{n=1}^{\infty} \mathbf{U}_n^I$ is convergent with probability 1. Note that the multigeneration offspring of every members of an arbitrary type $i \in I^c$ vanish at most in p steps. This means that $\mathbf{U}_n^{I^c} =_{\mathcal{D}} \mathbf{V}_n^{I^c}(1) = \mathbf{0}$ for any $n \geq p+1$. Let us recall that $\mathbf{Y}_n$ denotes that number of the nth generation offspring of the initial population $\mathbf{X}_0$. Since this process dies out in finitely steps with probability 1 in case of any initial distribution, we obtain the almost sure convergence $\mathbf{Y}_n \to \mathbf{0}$, $n \to \infty$. Then, by using (3.9) we get that

$$\mathbf{X}_n = \mathbf{Y}_n + \mathbf{Z}_n = \mathbf{Y}_n + \mathbf{Z}_n^I + \mathbf{Z}_n^{I^c} \xrightarrow{\mathcal{D}} \mathbf{0} + \sum_{k=1}^{\infty} \mathbf{U}_k^I + \sum_{k=1}^{p} \mathbf{U}_k^{I^c}, \qquad n \to \infty,$$

where the limit variable is finite with probability 1, and its law does not depend on the initial distribution. This convergence implies that the law of the limit variable is a stationary distribution for the Galton–Watson process $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, in case of an arbitrary set I.

We prove the contrary direction of the theorem by contradiction. For this goal, assume that the process has a stationary distribution π, and there exist states j_0, j and integers $m_0 \geq 0$, $m \geq 1$ such that $M_{j_0,j}^{(m_0)} > 0$, $M_{j,j}^{(m)} > 0$

and $\sum_{k=1}^{\infty} \log k P(\eta_{j_0}(1) = k) = \infty$. Since the second inequality implies that $M_{j,j}^{(nm)} > 0$ for any positive integer n, we can assume without the loss of generality that $m > m_0$. It also follows that $E \log(\eta_{j_0}(1) + 1) = \infty$. Let us fix the states j_0, j and the values m_0, m for the rest of the proof.

For an arbitrary positive integer n the members of the nmth generation of the process can be divided into two groups. Some of the members are mth generation offspring of the population $\mathbf{X}_{(n-1)m}$, and the others are members of the innovation $\boldsymbol{\eta}(nm)$ or multigeneration offspring of $\boldsymbol{\eta}((n-1)m), \ldots, \boldsymbol{\eta}(nm-1)$. By the branching mechanism the distribution of the number of the members in the second group is the same as the distribution of $\mathbf{Z}_m$. Also, the number of the mth generation offspring of an arbitrary member of type i in the $(n-1)m$th generation has the same distribution as the conditional law of $\mathbf{Y}_m$ with respect to $\{\mathbf{X}_0 = \mathbf{e}_i\}$. Furthermore, these variables are independent of each other and of $\mathbf{X}_{(n-1)m}$. Now, consider random vectors $\boldsymbol{\xi}_i^{(m)}(n,k), \boldsymbol{\eta}^{(m)}(n)$, $i = 1, \ldots, p$, $n, k = 1, 2, \ldots$ being independent of each other and of $\mathbf{X}_0$ such that $\boldsymbol{\xi}_i^{(m)}(n,k)$ has the same law as the conditional distribution of $\mathbf{Y}_m$ under $\{\mathbf{X}_0 = \mathbf{e}_i\}$, and $\boldsymbol{\eta}^{(m)}(n)$ has the same distribution as $\mathbf{Z}_m$, respectively. Then, we have

$$\mathbf{X}_{nm} \overset{\mathcal{D}}{=} \sum_{i=1}^{p} \sum_{k=1}^{X_{(n-1)m,i}} \boldsymbol{\xi}_i^{(m)}(n,k) + \boldsymbol{\eta}^{(m)}(n), \qquad n = 1, 2, \ldots \tag{3.10}$$

(We note that this equation can be proved by standard calculations too, by using generating functions). In the following we assume that equation (3.10) holds in almost sure sense for every n. We can do so without the loss of generality, because under this assumption the distribution of the process $\mathbf{X}_{nm}$, $n \in \mathbb{Z}_+$, does not change.

We recall that we fixed a state j earlier in the proof. Let $\mathbf{X}_0^*, \mathbf{X}_1^*, \ldots$ stands for the p-type Galton–Watson process with immigration corresponding to the offspring and innovation vectors

$$\boldsymbol{\xi}_i^*(n,k) = \mathbb{1}_{\{i=j\}} \boldsymbol{\xi}_i^{(m)}(n,k), \quad \boldsymbol{\eta}^*(n) = \boldsymbol{\eta}^{(m)}(n), \quad i = 1, \ldots, p, \quad n, k = 1, 2, \ldots,$$

with the initial value $\mathbf{X}_0^* = \mathbf{X}_0$. Then, we have $\boldsymbol{\xi}_i^*(n,k) \le \boldsymbol{\xi}_i^{(m)}(n,k)$ and $\boldsymbol{\eta}^*(n) \le \boldsymbol{\eta}^{(m)}(n)$ for every i, n and k with probability 1, and $(\mathbf{X}_0, \mathbf{X}_0^*)$ is independent of all of these offpsring and innovation variables. Since π is a stationary distribution of the Markov chain $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, it is a stationary distribution of the subsequence $\mathbf{X}_{nm}$, $n \in \mathbb{Z}_+$. Then, Proposition 3.3 implies that the process $\mathbf{X}_n^*$, $n = 0, 1, \ldots$, has a stationary distribution π^* too. Observe that we have

$$X_{n,j}^* = \sum_{k=1}^{X_{n-1,j}^*} \xi_{j,j}^*(n,k) + \eta_j^*(n), \qquad n = 1, 2, \ldots, \tag{3.11}$$

meaning that $X_{n,j}^*$, $n \in \mathbb{Z}_+$, is a single-type Galton–Watson process with immigration. Then, the jth marginal of the measure π^* is a stationary distribution for this Markov chain. However, in the next step we show that $X_{n,j}^*$, $n \in \mathbb{Z}_+$, does not have any stationary distribution, which leads to a contradiction. This proves that the moment condition of the theorem is necessary for the existence of a stationary distribution of the original process $\mathbf{X}_n$, $n \in \mathbb{Z}_+$.

To show that the process in (3.11) does not have any stationary distribution note that $\eta_j^*(1)$ has the same distribution as the jth component $Z_{m,j}$ of the vector $\mathbf{Z}_m$. Furthermore, $Z_{m,j}$ denotes the total number of those elements in the population $\mathbf{X}_m$ which are both of type j and members of the innovation $\boldsymbol{\eta}(m)$ or multigeneration offspring of $\boldsymbol{\eta}(1), \ldots, \boldsymbol{\eta}(m-1)$. Define $\eta = \eta_{j_0}(m - m_0)$, which is the number of those members in the innovation $\boldsymbol{\eta}(m - m_0)$ which are of type j_0. Let V stand for the number of those members in $\mathbf{X}_m$ which are of type j and m_0th generation offspring of the population η. Then, we have $Z_{m,j} \ge V$ with probability 1. Note that the conditional distribution of $Y_{m_0,j}$ with respect to $\{\mathbf{X}_0 = \mathbf{e}_{j_0}\}$ is the distribution of the number of those m_0th generation offspring of a single member of type j_0 which are of type j. Consider random variables $\xi_1, \xi_2, \ldots$

being independent of each other and of η and having identical law $\mathcal{L}(Y_{m_0,j}|\mathbf{X}_0 = \mathbf{e}_{j_0})$. Then, by the branching mechanism the variable V has the same distribution as the sum $\sum_{k=1}^{\eta} \xi_k$. Note that $E\xi_1 = M_{j_0,j}^{(m_0)} \in (0,\infty)$ and $E\log(\eta+1) = \infty$ by assumption. From these Proposition 3.1 implies that

$$E\log\big(\eta_j^*(1)+1\big) = E\log\big(Z_{m,j}+1\big) \geq E\log\big(V+1\big) = E\log\left(\sum_{k=1}^{\eta}\xi_k+1\right) = \infty.$$

Now, let us return to equation (3.11). Since the offspring variable $\xi_{j,j}^*(1,1)$ has the same distribution as $\xi_{j,j}^{(m)}(1,1)$, it follows that $E\xi_{j,j}^*(1,1) = M_{j,j}^{(m)} > 0$. Then, by Remark 3.4 the single-type Galton–Watson process $X_{n,j}^*$, $n \in \mathbb{Z}_+$, can not have any stationary distribution. This argument completes the proof of the theorem. □

Proof of Theorem 2.4. First, we show that if either (i) or (ii) is satisfied then $\mathcal{C}$ is a subset of a lower dimensional affine subspace $\mathcal{S}$ of $\mathbb{R}^p$. Assume that type j is minor for some $j = 1,\dots,p$. By using the notations introduced in the proof of Theorem 2.1, the jth component of $\mathbf{Z}_n$ vanishes with probability 1 for every positive integer n. Since $\mathcal{C}$ is defined as the union of the ranges of the variables $\mathbf{Z}_n$, $n \geq n^*$, the class $\mathcal{C}$ is a subset of the linear subspace $\mathcal{S}$ defined by the the equation $\mathbf{e}_j^\top \mathbf{v} = 0$, $\mathbf{v} \in \mathbb{R}^p$.

Now assume that (ii) holds, and consider an arbitrary $\mathbf{x}' \in \mathcal{C}$. Since $\mathcal{C}$ is a communication class, there exists a state $\mathbf{x} \in \mathcal{C}$ such that $\mathbf{x}'$ is accessible from $\mathbf{x}$ in one step. Working on the event $\{\mathbf{X}_0 = \mathbf{x}\}$ we get the equation

$$\mathbf{c}^\top \mathbf{X}_1 = \sum_{i=1}^{p}\sum_{k=1}^{x_i} \mathbf{c}^\top \boldsymbol{\xi}_i(1,k) + \mathbf{c}^\top \boldsymbol{\eta}(1) = 0 + \mathbf{c}^\top \boldsymbol{\eta}(1).$$

Because $P_{\mathbf{x}}(\mathbf{X}_1 = \mathbf{x}') > 0$ and $\mathbf{c}^\top \boldsymbol{\eta}(1)$ is degenerate by assumption, the vector $\mathbf{x}'$ is an element of the affine subspace $\mathcal{S}$ defined by the equation $\mathbf{c}^\top \mathbf{v} = \mathbf{c}^\top \boldsymbol{\eta}(1)$, $\mathbf{v} \in \mathbb{R}^p$.

For the contrary direction let $\mathcal{S} \subsetneq \mathbb{R}^p$ denote the affine subspace generated by $\mathcal{C}$, and assume that none of the types is minor. Consider an arbitrary state $\mathbf{x}^* \in \mathcal{C}$, and fix a vector $\mathbf{y}^* \in \mathbb{Z}_+^p$ such that $P(\boldsymbol{\eta}(1) = \mathbf{y}^*) > 0$. Since the set $\mathcal{V} = \mathcal{S} - \mathbf{x}^*$ is a linear subspace of $\mathbb{R}^p$ with dimension less than p, the orthogonal complement $\mathcal{V}^\perp$ of $\mathcal{V}$ is a non-trivial linear subspace of $\mathbb{R}^p$, and we have $\mathbf{c}^\top \mathbf{x} = \mathbf{c}^\top \mathbf{x}^*$ for every $\mathbf{c} \in \mathcal{V}^\perp$ and $\mathbf{x} \in \mathcal{S}$.

Consider an arbitrary state $\mathbf{x} \in \mathcal{C}$ and an arbitrary vector $\mathbf{c} \in \mathcal{V}^\perp$, and work on the event $\{\mathbf{X}_0 = \mathbf{x}\}$. The communication class $\mathcal{C}$ is closed, which implies that the variable $\mathbf{X}_1$ lies in $\mathcal{S}$ with probability 1, and hence, we obtain the equation

$$\mathbf{c}^\top \mathbf{x}^* = \mathbf{c}^\top \mathbf{X}_1 = \sum_{k=1}^{x_1} \mathbf{c}^\top \boldsymbol{\xi}_1(1,k) + \cdots + \sum_{k=1}^{x_p} \mathbf{c}^\top \boldsymbol{\xi}_p(1,k) + \mathbf{c}^\top \boldsymbol{\eta}(1). \tag{3.12}$$

Note that the left side of this equation is deterministic, and consider any type $i = 1,\dots,p$. Since type i in not minor by assumption, there exists a state $\mathbf{x} \in \mathcal{C}$ such that $x_i \neq 0$. If we apply equation (3.12) for this given state $\mathbf{x}$ then we have multiple independent terms on the right side including $\mathbf{c}^\top \boldsymbol{\xi}_i(1,1)$ and $\mathbf{c}^\top \boldsymbol{\eta}(1)$. Since the sum is deterministic, these terms are degenerate variables. This means that for an arbitrary vector $\mathbf{x} \in \mathcal{C}$ all of the terms on the right side of (3.12) are degenerate.

Let us introduce the variable

$$\mathbf{S}(\mathbf{x}) = \sum_{k=1}^{x_1} \boldsymbol{\xi}_1(1,k) + \cdots + \sum_{k=1}^{x_p} \boldsymbol{\xi}_p(1,k), \tag{3.13}$$

which is the number of the 1st generation offspring of the initial population $\mathbf{X}_0$ under the condition $\{\mathbf{X}_0 = \mathbf{x}\}$. Note that $E\mathbf{S}(\mathbf{x}) = \mathbf{M}^\top \mathbf{x}$. Then, we have

$$\mathbf{c}^\top \mathbf{x}^* = \mathbf{c}^\top \mathbf{S}(\mathbf{x}) + \mathbf{c}^\top \boldsymbol{\eta}(1) = E_{\mathbf{x}}\big(\mathbf{c}^\top \mathbf{S}(\mathbf{x})\big) + \mathbf{c}^\top \mathbf{y}^* = (\mathbf{M}\mathbf{c})^\top \mathbf{x} + \mathbf{c}^\top \mathbf{y}^* \tag{3.14}$$

with probability 1 for every state $\mathbf{x} \in \mathcal{C}$. Since $\mathcal{S}$ is the affine subspace generated by the set $\mathcal{C}$, equation (3.14) is valid for any $\mathbf{x} \in \mathcal{S}$, as well.

Consider an arbitrary vector $\mathbf{v} \in \mathcal{V}$, and note that both $\mathbf{v} + \mathbf{x}^*$ and $\mathbf{x}^*$ are elements of $\mathcal{S}$. Then, from equation (3.14) it follows that

$$(\mathbf{M}\mathbf{c})^\top \mathbf{v} = (\mathbf{M}\mathbf{c})^\top (\mathbf{v} + \mathbf{x}^*) - (\mathbf{M}\mathbf{c})^\top \mathbf{x}^* = \mathbf{c}^\top (\mathbf{x}^* - \mathbf{y}^*) - \mathbf{c}^\top (\mathbf{x}^* - \mathbf{y}^*) = 0.$$

This implies that $\mathbf{M}\mathbf{c} \in \mathcal{V}^\perp$ for any $\mathbf{c} \in \mathcal{V}^\perp$, and hence, we have $\psi(\mathcal{V}^\perp) \subseteq \mathcal{V}^\perp$ with the linear function $\psi : \mathbb{R}^p \to \mathbb{R}^p$, $\psi(\mathbf{c}) = \mathbf{M}\mathbf{c}$. Because the process $\mathbf{X}_n$, $n = 0, 1, \ldots,$ reaches the class $\mathcal{C}$ in finitely many steps almost surely in case of any initial state, there exists a state $\mathbf{z} \in \mathbb{Z}_+^p$, $\mathbf{z} \notin \mathcal{S}$, such that the subspace $\mathcal{S}$ is accessible from $\mathbf{z}$ in one step. Since under the event $\{\mathbf{X}_0 = \mathbf{z}\}$ the variable $\mathbf{X}_1$ is an element of $\mathcal{S}$ with positive probability, the equation

$$\mathbf{c}^\top \mathbf{x}^* = \mathbf{c}^\top \mathbf{X}_1 = \mathbf{c}^\top \mathbf{S}(\mathbf{z}) + \mathbf{c}^\top \boldsymbol{\eta}(1) = E_{\mathbf{z}}\big(\mathbf{c}^\top \mathbf{S}(\mathbf{z})\big) + \mathbf{c}^\top \mathbf{y}^* = (\mathbf{M}\mathbf{c})^\top \mathbf{z} + \mathbf{c}^\top \mathbf{y}^* \tag{3.15}$$

holds with positive probability in case of any $\mathbf{c} \in \mathcal{V}^\perp$. As a consequence, we get that $\mathbf{c}^\top (\mathbf{x}^* - \mathbf{y}^*) = (\mathbf{M}\mathbf{c})^\top \mathbf{z}$. Let us consider the orthogonal decomposition $\mathbf{z} = \mathbf{x} + \mathbf{x}^\perp$ where $\mathbf{x} \in \mathcal{S}$ and $\mathbf{x}^\perp \in \mathcal{V}^\perp$, $\mathbf{x}^\perp \neq \mathbf{0}$. From (3.14) we obtain that

$$\mathbf{c}^\top (\mathbf{x}^* - \mathbf{y}^*) = (\mathbf{M}\mathbf{c})^\top \mathbf{z} = (\mathbf{M}\mathbf{c})^\top \mathbf{x} + (\mathbf{M}\mathbf{c})^\top \mathbf{x}^\perp = \mathbf{c}^\top (\mathbf{x}^* - \mathbf{y}^*) + \psi(\mathbf{c})^\top \mathbf{x}^\perp,$$

and hence, $\psi(\mathbf{c})^\top \mathbf{x}^\perp = 0$ for any vector $\mathbf{c} \in \mathcal{V}^\perp$. That is, $\psi(\mathcal{V}^\perp) \perp \mathbf{x}^\perp \in \mathcal{V}^\perp$ implying that $\psi(\mathcal{V}^\perp) \subsetneq \mathcal{V}^\perp$. This means that ψ is not a full rank linear transformation, and there exists a vector $\mathbf{c}^* \in \mathcal{V}^\perp$ such that $\mathbf{M}\mathbf{c}^* = \psi(\mathbf{c}^*) = \mathbf{0}$. Since the variables $\mathbf{c}^\top \boldsymbol{\xi}_i(1,1)$, $i = 1, \ldots, p$, are deterministic in case of any $\mathbf{c} \in \mathcal{V}^\perp$, we get that

$$(\mathbf{c}^*)^\top \boldsymbol{\xi}_i(1,1) = E\big((\mathbf{c}^*)^\top \boldsymbol{\xi}_i(1,1)\big) = (\mathbf{c}^*)^\top E\boldsymbol{\xi}_i(1,1) = (\mathbf{M}\mathbf{c}^*)^\top \mathbf{e}_i = 0, \qquad i = 1, \ldots, p,$$

and the proof is complete. □

Proposition 3.5. *Let $\mathbf{A} \in \mathbb{R}^{p\times p}$ be a matrix having only nonnegative entries. If $\varrho(\mathbf{A}) < 1$, then there exist a constant $\lambda \in (0,1)$ and a vector $\mathbf{v} \in \mathbb{R}^p$ such that all components of $\mathbf{v}$ are strictly positive and $\mathbf{A}\mathbf{v} \leq \lambda\mathbf{v}$.*

Proof. Since the eigenvalues are continuous functions of the matrix entries, there exists an $\varepsilon > 0$ such that $\varrho(\mathbf{A} + \varepsilon) < 1$. Then, the Perron–Frobenius theorem implies that the positive matrix $\mathbf{A} + \varepsilon$ has an eigenvector $\mathbf{v}$ with eigenvalue $\lambda = \varrho(\mathbf{A} + \varepsilon)$ such that all components of $\mathbf{v}$ are strictly positive. With this vector we get the inequality $\mathbf{A}\mathbf{v} \leq (\mathbf{A} + \varepsilon)\mathbf{v} = \lambda\mathbf{v}$. □

Proof of Theorem 2.5. First, assume that the stationary distribution π has finite moment of order α, and consider an arbitrary type i that is not minor. Then, there exists a state $\mathbf{x} \in \mathcal{C}$ whose ith component is not zero. If the initial distribution of the process is set to the stationary distribution π, then

$$\infty > E\|\mathbf{X}_1\|^\alpha \geq E_{\mathbf{x}}\|\mathbf{X}_1\|^\alpha P(\mathbf{X}_0 = \mathbf{x}) \geq E\big\|\boldsymbol{\xi}_i(1,1)\big\|^\alpha \pi\big(\{\mathbf{x}\}\big),$$

proving that $E\|\boldsymbol{\xi}_i(1,1)\|^\alpha$ is finite. Similarly, $E\|\boldsymbol{\eta}(1)\|^\alpha \leq E\|\mathbf{X}_1\|^\alpha < \infty$.

For the contrary direction assume that (ii) holds, and consider the constant $\lambda \in (0,1)$ and the vector $\mathbf{v} \in \mathbb{Z}_+^p$ of Proposition 3.5 with $\mathbf{A} = \mathbf{M}$. Since the proposition remains true if we multiply $\mathbf{v}$ with a positive number, we can assume that the components of $\mathbf{v}$ are larger than 1. Also, introduce the function $V(\mathbf{x}) = (\mathbf{v}^\top \mathbf{x})^\alpha + 1$, $\mathbf{x} \in \mathbb{Z}_+^p$, and the set

$$\mathcal{F}_{\alpha,\mathbf{v}} = \big\{ f : \mathbb{Z}_+^p \to \mathbb{R} : |f(\mathbf{x})| \le V(\mathbf{x}), \mathbf{x} \in \mathbb{Z}_+^p \big\}.$$

Our goal is to prove that

$$E_{\mathbf{x}} V(\mathbf{X}_1) - V(\mathbf{x}) \le -c_1 V(\mathbf{x}) + c_2 \mathbb{1}_{\mathcal{Z}'}(\mathbf{x}), \qquad \mathbf{x} \in \mathcal{Z}, \tag{3.16}$$

where $c_1 > 0$ and c_2 are suitable real values, and $\mathbb{1}_{\mathcal{Z}'}$ is the indicator function of a suitable finite set $\mathcal{Z}' \subseteq \mathcal{Z}$. From Theorem 2.1 it follows that the process $\mathbf{X}_n$, $n \in \mathbb{Z}_+$, is *ψ-irreducible* and *aperiodic* in the sense of Meyn and Tweedie [10] on the reduced state space $\mathcal{Z}$, and their Proposition 5.5.5 implies that the finite set $\mathcal{Z}'$ is *petite*. This means that if we can prove the Foster–Lyapunov type criteria (3.16) for every state $\mathbf{x} \in \mathcal{Z}$ then Theorem 15.0.1 of [10] implies the inequality

$$\sum_{n=0}^{\infty} r^n \sup_{f \in \mathcal{F}_{\alpha,\mathbf{v}}} \left| E_{\mathbf{x}} f(\mathbf{X}_n) - \int_{\mathbb{Z}_+^p} f(\mathbf{y}) \pi(d\mathbf{y}) \right| \le R V(\mathbf{x}), \qquad \mathbf{x} \in \mathcal{Z},$$

with some finite constants $r > 1$ and $R > 0$. Since the components of $\mathbf{v} = (v_1, \dots, v_p)^\top$ are larger than 1, we obtain that

$$\|\mathbf{x}\|^\alpha + 1 \le V(\mathbf{x}) \le \|\mathbf{v}\|^\alpha \big(\|\mathbf{x}\|^\alpha + 1 \big), \qquad \mathbf{x} \in \mathbb{Z}_+^p.$$

These imply that the function $\mathbf{x} \mapsto \|\mathbf{x}\|^\alpha$ is an element of the set $\mathcal{F}_{\alpha,\mathbf{v}}$, and it has finite integral with respect to the measure π. Furthermore, we have $\mathcal{F}_\alpha \subseteq \mathcal{F}_{\alpha,\mathbf{v}}$, leading to the inequalities

$$\sum_{n=0}^{\infty} r^n \sup_{f \in \mathcal{F}_{\alpha}} \left| E_{\mathbf{x}} f(\mathbf{X}_n) - \int_{\mathbb{Z}_+^p} f(\mathbf{y}) \pi(d\mathbf{y}) \right| \le R V(\mathbf{x}) \le R \|\mathbf{v}\|^\alpha \big(\|\mathbf{x}\|^\alpha + 1 \big), \quad \mathbf{x} \in \mathbb{Z}_+^p.$$

That is, the theorem is verified with $a_1 = r$ and $a_2 = R\|\mathbf{v}\|^\alpha$.

In the rest of the proof we prove formula (3.16). If $\alpha \le 1$ then the function $t \mapsto t^\alpha$ is nonnegative and concave on the positive halfline. This implies that $(s+t)^\alpha \le s^\alpha + t^\alpha$ for any $s, t \ge 0$. By using formula (3.13) and Jensen's inequality we get that

$$\begin{aligned} &E_{\mathbf{x}} \big(\mathbf{v}^\top \mathbf{X}_1 \big)^\alpha \le E \big(\mathbf{v}^\top \mathbf{S}(\mathbf{x}) \big)^\alpha + E \big(\mathbf{v}^\top \boldsymbol{\eta}(1) \big)^\alpha \le \big(\mathbf{v}^\top E \mathbf{S}(\mathbf{x}) \big)^\alpha + E \big(\|\mathbf{v}\| \|\boldsymbol{\eta}(1)\| \big)^\alpha \\ &= \big(\mathbf{v}^\top \mathbf{M}^\top \mathbf{x} \big)^\alpha + \|\mathbf{v}\|^\alpha E \|\boldsymbol{\eta}(1)\|^\alpha \le \big(\lambda \mathbf{v}^\top \mathbf{x} \big)^\alpha + \|\mathbf{v}\|^\alpha E \|\boldsymbol{\eta}(1)\|^\alpha. \end{aligned}$$

Consider an arbitrary constant $c_1 \in (0, 1 - \lambda^\alpha)$ and the finite set

$$\mathcal{Z}' = \left\{ \mathbf{x} \in \mathcal{Z} : V(\mathbf{x}) < \frac{\|\mathbf{v}\|^\alpha E \|\boldsymbol{\eta}(1)\|^\alpha + 1}{1 - \lambda^\alpha - c_1} \right\}.$$

Then, for any $\mathbf{x} \in \mathcal{Z} \setminus \mathcal{Z}'$ we have

$$E_{\mathbf{x}} V(\mathbf{X}_1) \le \lambda^\alpha V(\mathbf{x}) + \|\mathbf{v}\|^\alpha E \|\boldsymbol{\eta}(1)\|^\alpha + 1 \le (1 - c_1) V(\mathbf{x}).$$

This implies inequality (3.16) with

$$c_2 = \max_{\mathbf{x}\in\mathcal{Z}'} \lambda^\alpha V(\mathbf{x}) + \|\mathbf{v}\|^\alpha E\|\boldsymbol{\eta}(1)\|^\alpha + 1 < \infty.$$

In the case $\alpha > 1$ let $\|\cdot\|_{L^\alpha}$ stand for the L^α-norm of random variables. Fix an arbitrary state $\mathbf{x}\in\mathcal{Z}$, and introduce the random vectors $\overline{\mathbf{X}}_1 = \mathbf{X}_1 - E_{\mathbf{x}}\mathbf{X}_1$,

$$\overline{\boldsymbol{\xi}}_i(n,k) = \boldsymbol{\xi}_i(n,k) - E\boldsymbol{\xi}_i(n,k), \quad \overline{\boldsymbol{\eta}}(n) = \boldsymbol{\eta}(n) - E\boldsymbol{\eta}(n), \quad i = 1,\dots,p,\ n,k = 1,2,\dots$$

Then, we have

$$\begin{aligned}\big(E_{\mathbf{x}}(\mathbf{v}^\top\mathbf{X}_1)^\alpha\big)^{1/\alpha} &= \big\|\mathbf{v}^\top\mathbf{X}_1\big\|_{L^\alpha} = \big\|\mathbf{v}^\top\overline{\mathbf{X}}_1 + E\big(\mathbf{v}^\top\mathbf{S}(\mathbf{x})\big) + E\big(\mathbf{v}^\top\boldsymbol{\eta}(1)\big)\big\|_{L^\alpha}\\ &\le \big\|\mathbf{v}^\top\overline{\mathbf{X}}_1\big\|_{L^\alpha} + \big\|\mathbf{v}^\top\mathbf{M}^\top\mathbf{x}\big\|_{L^\alpha} + \big\|\mathbf{v}^\top E\boldsymbol{\eta}(1)\big\|_{L^\alpha} \le \big\|\mathbf{v}^\top\overline{\mathbf{X}}_1\big\|_{L^\alpha} + \lambda\mathbf{v}^\top\mathbf{x} + \mathbf{v}^\top E\boldsymbol{\eta}(1).\end{aligned}$$

By using the Marcinkiewicz–Zygmund inequality (see Thm. 13 of Marcinkiewicz and Zygmund [13] or Theorem 10.3.2 of Chow and Teicher [14]) we get that

$$\big\|\mathbf{v}^\top\overline{\mathbf{X}}_1\big\|_{L^\alpha}^\alpha = E\bigg|\sum_{i=1}^p\sum_{k=1}^{x_i}\mathbf{v}^\top\overline{\boldsymbol{\xi}}_i(1,k) + \mathbf{v}^\top\overline{\boldsymbol{\eta}}(1)\bigg|^\alpha \le CE\bigg[\sum_{i=1}^p\sum_{k=1}^{x_i}\big|\mathbf{v}^\top\overline{\boldsymbol{\xi}}_i(1,k)\big|^2 + \big|\mathbf{v}^\top\overline{\boldsymbol{\eta}}(1)\big|^2\bigg]^{\alpha/2},$$

where C is a suitable positive constant depending only on α. Let us note that for arbitrary nonnegative real numbers $a_1,\dots,a_n$ we have

$$\big(a_1+\cdots+a_n\big)^{\alpha/2} \le n^{\beta-1}\big(a_1^{\alpha/2}+\cdots+a_n^{\alpha/2}\big),$$

where the value β also depends only on α. This inequality follows with $\beta = 1$ in the case $\alpha \le 2$ from the fact that the function $t\mapsto t^{\alpha/2}$ is concave on the positive halfline, and with $\beta = \alpha/2$ in the case $\alpha > 2$ from the power mean inequality. Then, we get that

$$\big\|\mathbf{v}^\top\overline{\mathbf{X}}_1\big\|_{L^\alpha}^\alpha \le C\big(\|\mathbf{x}\|+1\big)^{\beta-1}\bigg[\sum_{i=1}^p\sum_{k=1}^{x_i}E\big|\mathbf{v}^\top\overline{\boldsymbol{\xi}}_i(1,k)\big|^\alpha + E\big|\mathbf{v}^\top\overline{\boldsymbol{\eta}}(1)\big|^\alpha\bigg] \le C\big(\|\mathbf{x}\|+1\big)^\beta b_\alpha,$$

where

$$b_\alpha = \max\Big\{E\big|\mathbf{v}^\top\overline{\boldsymbol{\xi}}_1(1,1)\big|^\alpha,\dots,E\big|\mathbf{v}^\top\overline{\boldsymbol{\xi}}_p(1,1)\big|^\alpha, E\big|\mathbf{v}^\top\overline{\boldsymbol{\eta}}(1)\big|^\alpha\Big\} < \infty.$$

Consider any constant $c_1\in(0,1-\lambda^\alpha)$. Since $(1-c_1)^{1/\alpha} > \lambda$ and $\beta/\alpha < 1$, it follows that

$$\begin{aligned}E_{\mathbf{x}}V(\mathbf{X}_1) = E_{\mathbf{x}}\big(\mathbf{v}^\top\mathbf{X}_1\big)^\alpha + 1 &\le \Big[(Cb_\alpha)^{1/\alpha}\big(\|\mathbf{x}\|+1\big)^{\beta/\alpha} + \lambda\mathbf{v}^\top\mathbf{x} + \mathbf{v}^\top E\boldsymbol{\eta}(1)\Big]^\alpha + 1\\ &\le \big[(1-c_1)^{1/\alpha}\mathbf{v}^\top\mathbf{x}\big]^\alpha \le (1-c_1)V(\mathbf{x}),\end{aligned}$$

where the second inequality holds for all except finitely many values $\mathbf{x}\in\mathcal{Z}$. Let $\mathcal{Z}'$ stand for the set of those states $\mathbf{x}\in\mathcal{Z}$ for which the second inequality does not hold. Then we obtain (3.16) in the case $\alpha > 1$ with

$$c_2 = \max_{\mathbf{x}\in\mathcal{Z}'}\Big[(Cb_\alpha)^{1/\alpha}\big(\|\mathbf{x}\|+1\big)^{\beta/\alpha} + \lambda\mathbf{v}^\top\mathbf{x} + \mathbf{v}^\top E\boldsymbol{\eta}(1)\Big]^\alpha + 1 < \infty.$$

If all offspring distributions have finite moment of order α, then one can show by similar calculations that inequality (3.16) holds for any $\mathbf{x} \in \mathbb{Z}_+^p$ with suitable constants c_1, c_2 and with a finite set $\mathcal{Z}' \subset \mathbb{Z}_+^p$. Then, again, Theorem 15.0.1 of Meyn and Tweedie [10] implies that (2.3) holds not only on the set $\mathcal{Z}$ but for every state $\mathbf{x} \in \mathbb{Z}_+^p$. This argument completes the proof of our last theorem. □

Acknowledgements

Project no TKP2021-NVA-09 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NVA funding scheme.

References

[1] K.B. Athreya and P. Ney, Branching Processes. Springer (1972).
[2] C.J. Mode, Multitype Branching Processes, Theory and Applications. American Elsevier Publishing Co. (1971).
[3] M.P. Quine, The multi-type Galton–Watson process with immigration. *J. Appl. Probab.* **7** (1970) 411–422.
[4] N. Kaplan, The multitype Galton–Watson process with immigration. *Ann. Probab.* **1** (1973) 947–953.
[5] M. Barczy, F.K. Nedényi and G. Pap, On aggregation of multitype Galton–Watson branching processes with immigration. *Mod. Stoch. Theory Appl.* **5** (2018) 53–79.
[6] P. Kevei and P. Wiandt, Moments of the stationary distribution of subcritical multitype Galton–Watson processes with immigration. *Stat. Probab. Lett.* **173** (2021) 109067.
[7] G. Pap and T.T. Szabó, Change detection in INAR(p) processes against various alternative hypotheses. *Comm. Statist. Theory Methods* **42** (2013) 1386–1405.
[8] F. Nedényi, Conditional least squares estimators for multitype Galton–Watson processes. *Acta Sci. Math. (Szeged)* **81** (2015) 325–348.
[9] J. Foster and P. Ney, Limit laws for decomposable critical branching processes. *Z. Wahrsch. Verw. Gebiete* **46** (1978) 13–43.
[10] S.P. Meyn and R.L. Tweedie, Markov Chains and Stochastic Stability. Cambridge University Press (2009).
[11] W. Feller, An Introduction to Probability Theory and its Applications. Vol. I. John Wiley & Sons (1968).
[12] J.H. Foster and J.A. Williamson, Limit theorems for the Galton-Watson process with time-dependent immigration. *Z. Wahrsch. Verw. Gebiete* **20** (1971) 227–235.
[13] J. Marcinkiewicz and A. Zygmund, Sur les foncions independantes. *Fund. Math.* **28** (1937) 60–90.
[14] Y.S. Chow and H. Teicher, Probability Theory. Independence, Interchangeability, Martingales. Springer-Verlag (1997).