Estimation of subcritical Galton Watson processes with correlated immigration

Yacouba Boubacar Maïnassara [1] & Landy Rabehasaina [2]

[1] *Laboratoire de Matériaux Céramiques et de Mathématiques, Université Polytechnique Hauts-de-France, INSA Hauts-de-France, Valenciennes, France, Email: Yacouba.BoubacarMainassara@uphf.fr*
[2] *Laboratoire de Mathématiques de Besançon, Université de Franche Comté, Besançon, France, Email: lrabehas@univ-fcomte.fr*

Résumé. Nous considérons un processus observé de Galton Watson $\{Y_n,\ n \in \mathbb{Z}\}$ avec immigration modélisé par un processus corrélé $\{\epsilon_n,\ n \in \mathbb{Z}\}$. Nous présentons des résultats d'estimation du taux de reproduction et l'espérance de l'immigration dans deux situations. La première est lorsque $\mathrm{Cov}(\epsilon_0, \epsilon_k) = 0$ pour k supérieur à un certain k_0: nous fournissons un estimateur et prouvons un résultat de normalité asymptotique. Dans un deuxième temps, nous considérons le cas où $\{\epsilon_n,\ n \in \mathbb{Z}\}$ a une structure de corrélation générale. Sous des hypothèses de mélange, nous déterminons un estimateur pour le taux de reproduction et nous montrons sa convergence en moyenne quadratique avec vitesse explicite. Lorsque le coefficient de mélange décroit suffisamment vite, un développement d'ordre 2 pour cet estimateur est établi. **Mots-clés.** Processus de Galton Watson, immigration, processus INAR.

Abstract. We consider an observed subcritical Galton Watson process $\{Y_n,\ n \in \mathbb{Z}\}$ with correlated stationary immigration process $\{\epsilon_n,\ n \in \mathbb{Z}\}$. Two situations are presented. The first one is when $\mathrm{Cov}(\epsilon_0, \epsilon_k) = 0$ for k larger than some k_0: a consistent estimator for the reproduction and mean immigration rates is given, and a CLT is proved. The second one is when $\{\epsilon_n,\ n \in \mathbb{Z}\}$ has general correlation structure: under mixing assumptions, we exhibit an estimator for the logarithm of the reproduction rate and we prove that it converges in quadratic mean with explicit speed. In addition, when the mixing coefficients decrease fast enough, we provide and prove a two terms expansion for the estimator.

Keywords. Galton Watson processes, immigration, INAR processes.

1 Introduction and model

Estimation of parameters in a Galton Watson process with immigration has a long history: we refer to the seminal paper Klimko and Nelson (1978) for laying the ground and expliciting conditional least square estimators for the expectation of the reproduction and immigration sequences in the subcritical case. A central limit theorem for these estimators was later proved in Venkataraman (1982), using a time series point of view. Note that the link between such processes and the so called integer valued times series $INAR(1)$ processes has been exploited, see e.g. Al-Osh and Alzaid (1987) which studied such a process with particular distributions for the reproduction and immigration sequences. A certain number of extensions for the model were later devised and studied. Wei and Winnicki (1990) considered the general critical and supercritical case and proved central limit theorems for (modified) weighted least square estimators. In Barczy et al (2021), the specific case where the immigration sequence has a regular variation distribution is considered, leading to asymptotic normality of the reproduction mean when the immigration mean is known. Generalization to two types processes have been recently investigated in Ispány et al (2014), Körmendi and Pap (2018), including estimations for the criticality parameter. Note that these references have one of the following constraint on the immigration process: first, some of them assume that its expectation is known (and appears in the expression of the estimator for the reproduction sequence); second, this immigration process is assumed to be a sequence of *independent* random variables, in addition to be identically distributed. Hence, we aim in this paper at considering a process which is stationary, but where some form of dependence is given. This particular feature appears not to have been studied in the literature.

We consider the following Galton Watson process with immigration as the stationary sequence $\{Y_n,\ n\in\mathbb{Z}\}$ satisfying

$$Y_{n+1}=\sum_{k=1}^{Y_n}\xi_{n+1,k}+\epsilon_{n+1},\quad n\in\mathbb{Z}, \tag{1}$$

for some sequences $\{\xi_{n,k},\ n\in\mathbb{Z},\ k\in\mathbb{N}\}$ and $\{\epsilon_n,\ n\in\mathbb{Z}\}$, named thereafter the *reproduction* sequence and the *immigration* sequence, which are such that

- $\{\xi_{n,k},\ n\in\mathbb{Z},\ k\in\mathbb{N}\}$ and $\{\epsilon_n,\ n\in\mathbb{Z}\}$ are independent sequences,
- the reproduction sequence $\{\xi_{n,k},\ n\in\mathbb{Z},\ k\in\mathbb{N}\}$ is an i.i.d. (doubly indexed) sequence, with distribution of a generic r.v. denoted by ξ,
- the immigration process $\{\epsilon_n,\ n\in\mathbb{Z}\}$ is stationary and ergodic, with distribution that of a generic r.v. denoted by ϵ,
- the moments $\lambda_0:=\mathbb{E}(\xi)$ and $m_0:=\mathbb{E}(\epsilon)$ of $\{\xi_{n,k},\ n\in\mathbb{Z},\ k\in\mathbb{N}\}$ and $\{\epsilon_n,\ n\in\mathbb{Z}\}$ are unknown.

The aim of the paper is to estimate the unknown reproduction rate λ_0 as well as the mean immigration m_0 from the observed sequence $\{Y_n,\ n\in\mathbb{Z}\}$. Furthermore, in order for the existence of the stationary process $\{Y_n,\ n\in\mathbb{Z}\}$ to exist, it is required that the subcritical case holds, namely that

$$\lambda_0<1. \tag{2}$$

To control the serial dependence of the stationary process $\{\epsilon_n,\ n\in\mathbb{Z}\}$, we introduce the strong mixing coefficients $\alpha_\epsilon(h)$ defined by

$$\alpha_\epsilon(h)=\sup_{A\in\mathcal{F}_{-\infty}^n,B\in\mathcal{F}_{n+h}^\infty}\left|\mathbb{P}(A\cap B)-\mathbb{P}(A)\mathbb{P}(B)\right|,$$

where $\mathcal{F}_{-\infty}^n=\sigma(\epsilon_u,u\le n)$ and $\mathcal{F}_{n+h}^\infty=\sigma(\epsilon_u,u\ge n+h)$. Note that $\alpha_\epsilon(h)$ does not depend on $n\in\mathbb{Z}$ thanks to the stationarity of $\{\epsilon_n,\ n\in\mathbb{Z}\}$.

To establish the asymptotic properties of the proposed estimator of λ_0, the following assumptions are required.

(A1): The mixing coefficient $\alpha_\epsilon(\cdot)$ of $\{\epsilon_n,\ n \in \mathbb{Z}\}$ verifies the following summability condition:

$$\sum_{h=0}^{\infty} \lambda_0^{-h} \alpha_\epsilon(h)^{1-2/\beta} < \infty \quad \text{for some} \quad \beta > 2.$$

We next make an integrability assumption on the moments and covariances of the immigration process $\{\epsilon_n,\ n \in \mathbb{Z}\}$ and reproduction sequence $\{\xi_{n,k},\ n \in \mathbb{Z},\ k \in \mathbb{N}\}$. We use $\|\cdot\|$ to denote the Euclidean norm of a vector and for any (potentially matrix valued) random variable X, we will set $\|Y\|_p^p := \mathbb{E}\|X\|^p$ its $\mathbb{L}^p$ norm, with $p \geq 1$.

(A2): The following moment conditions hold:

$$\|\epsilon\|_{2\beta} = \left[\mathbb{E}\,|\epsilon|^{2\beta}\right]^{1/(2\beta)} < \infty \quad \text{and} \quad \|\xi\|_{2\beta} = \left[\mathbb{E}\,|\xi|^{2\beta}\right]^{1/(2\beta)} < \infty.$$

(A3): The covariance of the immigration process $\nu_h := \mathrm{Cov}(\epsilon_0, \epsilon_h)$ verifies:

$$\sum_{h=0}^{\infty} h\lambda_0^{-h} \left|\nu_{h+1}\right| < \infty.$$

Let us then define for all $n \in \mathbb{N}$ and $k \geq 0$

$$\bar{Y}_n := \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \bar{Y}_{k+1,n} := \frac{1}{n}\sum_{i=1}^{n} Y_i Y_{i+k+1}.$$

2 Ultimately uncorrelated immigration

We suppose in this section that the immigration is no more correlated from an instant k_0, i.e. that the following assumption holds

$(\mathbf{A3})_\mathbf{1}$: There exists a known $k_0 \in \mathbb{N} \setminus \{0\}$ such that $\nu_k = 0$ for all $k \geq k_0$.

The main results are given as follows.

Proposition 1 (Consistency)**.** Let us suppose that (**A1**), (**A2**), (**A3**) and $(\mathbf{A3})_\mathbf{1}$ hold. Then the following estimators

$$\hat{R}_{k_0,n} := \frac{\left[\bar{Y}_n\right]^2 - \bar{Y}_{k_0,n}}{\left[\bar{Y}_n\right]^2 - \bar{Y}_{k_0-1,n}}, \quad \hat{M}_{k_0,n} := \bar{Y}_n(1 - \hat{R}_{k_0,n}) \tag{3}$$

converge a.s. towards the unknown parameters λ_0 and m_0 as $n \to \infty$.

Theorem 2 (Asymptotic normality)**.** Under the previous assumptions, the following CLT holds:

$$\sqrt{n}[\hat{R}_{k_0,n} - \lambda_0, \hat{M}_{k_0,n} - m_0]' \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \Omega_{k_0}), \tag{4}$$

where Ω_{k_0} is explicit.

3 General correlated immigration

In this case, additional Assumptions are required.

$(\mathbf{A3})_{\mathbf{2}}$: The mixing coefficient $\alpha_\epsilon(\cdot)$ has exponential decrease.

$(\mathbf{A4})$: The unknown parameters λ_0, m_0, $V_{0,\xi}$ and $V_{0,\epsilon}$ (reminding that the two latter are the variances of ξ and ϵ) belong to $\Theta := [\lambda_-, \lambda_+] \times \Theta_m \times \Theta_{V_\xi} \times \Theta_{V_\epsilon}$ for some known respective intervals $[\lambda_-, \lambda_+]$ and compact intervals Θ_m, Θ_{V_ξ} and Θ_{V_ϵ} included in $[0, \infty)$, with $0 < \lambda_- < \lambda_+ < 1$, and the generating function $f_\nu : x \in [0,1] \mapsto \sum_{s=0}^{\infty} x^s \nu_{s+1}$ belongs to some known class of function $\mathcal{F}$.

$(\mathbf{A5})$: There exists a known quantity $K_m > 0$ such that

$$\inf_{(\lambda,\mu,V_\xi,V_\epsilon)\in\Theta,\ f\in\mathcal{F}} |\Xi(\lambda, \mu, V_\xi, V_\epsilon, f)| \geq K_m \tag{5}$$

where Ξ is the function defined on $\Theta \times \mathcal{F}$ by

$$\Xi(\lambda, \mu, V_\xi, V_\epsilon, f) := -\frac{V_\xi m_0 \lambda}{(1-\lambda^2)(1+\lambda)} - \frac{V_\epsilon \lambda}{1-\lambda^2} - \frac{\lambda^2}{1-\lambda^2} f(\lambda) - \frac{1}{1-\lambda^2} f(\lambda^{-1}) \tag{6}$$

where $(\lambda, \mu, V_\xi, V_\epsilon, f) \in \Theta \times \mathcal{F}$.

$(\mathbf{A6})$: There exists a known constant $C_Y > 0$ such that the k-th moments $\|Y_0\|_k$ of the stationary process, $k = 1, 2$ are less than C_Y.

We give a few comments on the two last assumptions. It may be proved that Assumption $(\mathbf{A6})$ holds in one of the following situations:

- When $\mathcal{F}$ is the set of power series with non negative coefficents, meaning that the immigration sequence $\{\epsilon,\ n \in \mathbb{Z}\}$ is positively correlated, i.e. $\nu_n \geq 0$ for all $n \geq 1$, in which case K_m is explicit. This is the case when one models the evolution of a disease with constant (increasing or decreasing) trend.
- When $\mathcal{F}$ is included in the set of functions bounded by some constant $M_{\mathcal{F}}$, in which case we may prove that K_m is explcit and positive if the bound $M_{\mathcal{F}}$ is small enough.

As to $(\mathbf{A7})$, it may be proved that an explicit C_Y may be obtained e.g. if we assume that $\mathbb{E}(\xi^2) \leq 1$.

In order to state the main result of this section, we first introduce some auxiliary functions that will enable us to construct the estimator for the unknown parameters λ_0 and m_0. We first let $\varpi : \mathbb{R} \longrightarrow \mathbb{R}$ be a twice differentiable function such that

$$\varpi(x) \quad \begin{cases} = 1, & x \in [1, +\infty), \\ = 0, & x \in (-\infty, 0], \\ \in [0,1], & x \in [0,1], \end{cases} \tag{7}$$

$$\varpi''(x) = \mathrm{o}(x), \quad x \to 0, \tag{8}$$

so that $\varpi'(x) = \mathrm{o}(x^2)$ and $\varpi(x) = \mathrm{o}(x^3)$ as $x \to 0$.

We also need a twice differentiable function $G : \mathbb{R} \longrightarrow \mathbb{R}$ with finite support such that $G(x) = 1$ for $x \in [0, \max(C_Y, C_Y^2)]$. We then let

$$\varpi_k : x \in \mathbb{R} \mapsto \varpi\left(\frac{2}{K_m \lambda_-^k} x\right), \quad k \in \mathbb{N}, \tag{9}$$

K_m being defined in (5), so that $\varpi_k(x)$ is equal to 1 on $\left[2^{-1} K_m \lambda_-^k, \infty\right)$, lies in $[0,1]$ on $\left[0, 2^{-1} K_m \lambda_-^k\right]$ and is 0 on $(-\infty, 0]$. We finally define for all $k \in \mathbb{N} \setminus \{0\}$

$$H_k : x \in \mathbb{R} \mapsto \frac{1}{k} \varpi_k(|x|) \ln |x|, \tag{10}$$

$$\psi_k : (a,b) \in \mathbb{R}^2 \mapsto G(a)G(b)H_k(a^2 - b). \tag{11}$$

The following estimator is then defined:

$$\hat{S}_n := \psi_{k_n}(\bar{Y}_n, \bar{Y}_{k_n+1,n}),$$

for a sequence $(k_n)_{n\in\mathbb{N}}$ which is adequately chosen later on. We first state a consistency result.

Theorem 3 (Quadratic convergence of estimator)**.** Let us suppose that (**A1**), (**A2**), (**A3**), $(\mathbf{A3})_{\mathbf{2}}$, (**A4**), (**A5**) and (**A6**) hold. Let us set $k_n := \lfloor c \ln n \rfloor$ where $c < -\frac{1}{2 \ln \lambda_-}$. The following convergence in quadratic mean holds:

$$\left|\left|\hat{S}_n - \ln \lambda_0\right|\right|_2 = \mathrm{O}\left(\frac{1}{\ln n}\right) \longrightarrow 0, \quad n \to \infty \tag{12}$$

so that, in particular, $e^{\hat{S}_n}$ converges in probability towards λ_0 as $n \to \infty$. Furthermore, the estimator defined by $\hat{N}_n := \bar{Y}_n\left(1 - e^{\hat{S}_n}\right)$ converges in probability towards m_0 as $n \to \infty$.

The previous result may be refined under slightly stronger assumptions as follows.

Theorem 4 (Expansion for $\hat{S}_n$)**.** Let us in addition assume here that the covariance of the immigration process satisfies $\nu_h = \mathrm{O}(\zeta^h)$ for some $\zeta < \lambda_-$ (ensuring that (**A2**) holds). Let us set $k_n := \lfloor c \ln n \rfloor$ where $c \in \left(-\frac{1}{2\ln \zeta}, -\frac{1}{2 \ln \lambda_-}\right)$. Then one has the two terms expansion

$$\hat{S}_n - \ln \lambda_0 = \frac{1}{k_n} \ln \left| \sum_{j=0}^{\infty} \lambda_0^{-j} \chi_j + \lambda_0 (C_1^2 - C_2) \right| + \frac{1}{\sqrt{n} k_n \lambda_0^{k_n}} Z_n \tag{13}$$

where Z_n satisfies $Z_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma)$, $n \to \infty$, for some explicit $\sigma > 0$.

References

Al-Osh, M. A. and Alzaid, A. A. (1987), First-order integer-valued autoregressive (INAR(1)) process, *Journal of Time Series Analysis*, 8 (3), pp.261–275.

Barczy, M., Basrak, B., Kevei, P., Pap, G. and Planinić, H. (2021), Statistical inference of subcritical strongly stationary Galton-Watson processes with regularly varying immigration, *Stochastic Processes and their Applications*, 132, pp.33–75.

Klimko, L.A. and Nelson, P.I. (1978), On conditional least squares estimation for stochastic processes, *Annals of Statistics*, 6(3), pp.629–642.

Körmendi, K. and Pap, G. (2018), Statistical inference of 2-type critical Galton–Watson processes with immigration, *Statistical Inference for Stochastic Processes*, 21, pp.169–190.

Ispány, M., Körmendi, K. and Pap, G. (2014), Asymptotic behavior of CLS estimators for 2-type doubly symmetric critical Galton–Watson processes with immigration, *Bernoulli*, 20(4), pp.2247–2277.

Venkataraman, K.N. (1982), A time series approach to the study of the simple subcritical Galton–Watson process with immigration, *Advances in Applied Probability*, 14(1), pp.1–20.

Wei, C. Z. and Winnicki, J. (1990), Estimation of the means in the branching process with immigration, *Annals of Statistics*, 18(4), pp.1757–1773.