Weighted Estimation and Tracking for Branching Processes with Immigration

Bernard Bercu

***Abstract*—For branching processes with immigration, we propose a new approach which allows us to consistently estimate the means m, λ, and the variances σ^2, b^2 of the offspring and immigration distributions, respectively. Generally, statistical results for branching processes are established under the well-known trichotomy $m < 1$, $m = 1$, and $m > 1$. For example, no parameters of the immigration distribution can be consistently estimated if $m > 1$. The purpose of this paper is to obtain, through the introduction of a suitable adaptive control, strongly consistent estimators for all the parameters m, λ, σ^2, and b^2 without any restriction on the range of m. Central limit theorems and laws of iterated logarithm are also provided.**

***Index Terms*—Adaptive control, branching processes, central limit theorem, law of iterated logarithm, least-squares estimation.**

I. INTRODUCTION

THE purpose of this paper is to expand the adaptive control theory to the branching processes framework. We shall investigate the statistical properties of the controlled Bienaymé–Galton–Watson process with immigration (BGWI) given, for all $n \geq 0$, by

$$X_{n+1} = \sum_{i=1}^{X_n+U_n} Y_{n,i} + I_{n+1}$$

where $(Y_{n,i})$ and (I_n) are two independent sequences of independednt identically distributed (i.i.d.) nonnegative, integer-valued random variables. The initial variables X_0 and U_0 are integer-valued square-integrable random variables which are independent of $(Y_{n,i})$ and (I_n). The distribution of $(Y_{n,i})$, with finite mean m and positive variance σ^2, is commonly called the offspring distribution. The distribution of (I_n), with finite mean λ and positive variance b^2, is known as the immigration distribution. We are interested in the estimation of all the parameters m, λ, σ^2, and b^2.

For the classical BGWI process without control, Heyde and Seneta [18], [19] were the first to provide estimation results for m and λ without imposing restrictive assumptions on $(Y_{n,i})$ and (I_n). However, they do not solve the problem of how to estimate m and λ if we do not know whether $m < 1$, $m = 1$, or $m > 1$. More recently, Wei and Winnicki [28], [29] achieved consistency results for m and λ without any restriction on the range of m. However, they proved that there is no consistent estimator for λ and b^2 if $m > 1$. Furthermore, in the case where $m = 1$, Pakes [24] discovered the dichotomy $2\lambda \leq \sigma^2$ and $2\lambda > \sigma^2$. The process (X_n) is null recurrent if $2\lambda \leq \sigma^2$, whereas it is transient if $2\lambda > \sigma^2$. In the case where $m = 1$ and $2\lambda > \sigma^2$, Winnicki [31] established that the only parameters which may have consistent estimators are the first four moments of the offspring distribution and the mean λ of the immigration. Consequently, there is no consistent estimator for b^2 if $m = 1$ and $2\lambda > \sigma^2$.

The purpose of this paper is to establish estimation results for m, λ, σ^2, and b^2, without imposing restrictive assumptions on the parameters, by introducing a suitable adaptive control in the classical BGWI process. This adaptive control (U_n) regulates the dynamic of the process (X_n). On the one hand, (U_n) generates offsprings if there is not enough "energy" in the process. On the other hand, (U_n) eliminates them if there is too much "energy" in the process. By the same token, (U_n) also forces (X_n) to track step by step a given reference trajectory (x_n). We shall denote by $\mathbb{F} = (\mathcal{F}_n)$ with $\mathcal{F}_n = \sigma\{X_0, U_0, Y_{k,i}, I_k$ with $1 \leq k \leq n, i \geq 1\}$ the natural filtration of the model.

This paper is organized as follows. Section II is devoted to the estimation of the offspring parameters for the BGW process without immigration. We shall show the strong consistency of weighted least squares estimators of m and σ^2. Moreover, for each estimator, a central limit theorem (CLT) and a law of iterated logarithm (LIL) are also provided. In Section III, we establish similar estimation results for all parameters m, λ, σ^2 and b^2 in the BGWI framework. A short conclusion is given in Section IV where we mention some possible practical applications. All technical proofs are postponed in the Appendexes. We refer the reader to Asmussen and Hering [1], Athreya and Ney [2], and Guttorp [16] for basic properties of branching processes, and also to Caines [7], Chen and Guo [8], and Duflo [12] for the main results of adaptive control theory.

II. BGW RESULTS

We consider first the BGW process without immigration by taking (I_n) identically null so that

$$X_{n+1} = \sum_{i=1}^{X_n+U_n} Y_{n,i}. \tag{II.1}$$

In all the sequels, we assume that (U_n) is a sequence of integer-valued random variables adapted to $\mathbb{F}$ such that, whatever the value of $n \in \mathbb{N}$, $X_n + U_n \geq 1$. We can rewrite (II.1) as the autoregressive form

$$X_n = ma_n + \epsilon_n, \quad a_n = X_{n-1} + U_{n-1} \tag{II.2}$$

Manuscript received November 30, 1998; revised January 10, 2000. Recommended by Associate Editor Q. Zhang.

The author is with the Laboratoire de Statistiques, Université de Paris-Sud, 91 405 Orsay Cedex, France (e-mail: Bernard.Bercu@math.u-psud.fr).

Publisher Item Identifier S 0018-9286(01)00293-8.

where $\epsilon_n = X_n - ma_n$. Then, (ϵ_n) is a martingale difference sequence adapted to $\mathbb{F}$ such that $E[\epsilon_n^2|\mathcal{F}_{n-1}] = \sigma^2 a_n$. In order to estimate m, we propose to make use of the weighted least squares (WLS) estimator $\hat{m}_n$ that minimizes the quadratic criterion

$$\Delta_n(m) = \frac{1}{2} \sum_{k=1}^{n} a_k^{-1} (X_k - ma_k)^2.$$

Consequently, we obviously have

$$\hat{m}_n = A_n^{-1} \sum_{k=1}^{n} X_k \quad \text{where} \quad A_n = \sum_{k=1}^{n} a_k. \tag{II.3}$$

In addition, we also estimate the variance σ^2 by

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^{n} a_k^{-1}(X_k - \hat{m}_n a_k)^2. \tag{II.4}$$

We shall now specify the crucial choice of the adaptive control (U_n). The goal of adaptive tracking is to find a sequence (U_n) which regulates the dynamic of the process (X_n) by forcing (X_n) to follow a given reference trajectory (x_n). We assume that (x_n) is a sequence of nonnegative integer-valued random variables such that x_n is $\mathcal{F}_{n-1}$ measurable. We have, from (II.2),

$$X_n - x_n = \pi_n + \epsilon_n$$

where $\pi_n = ma_n - x_n$. If the parameter m was known, we would choose U_n such that π_{n+1} be as close as possible to zero, i.e., $U_n = P(m^{-1}x_{n+1}) - X_n$, where P denotes the projection operator on $\mathbb{N}$. Therefore, we propose to make use of the adaptive tracking control

$$U_n = \begin{cases} 1 - X_n, & \text{if } P(\hat{m}_n^{-1}x_{n+1}) = 0 \\ P(\hat{m}_n^{-1}x_{n+1}) - X_n, & \text{otherwise.} \end{cases} \tag{II.5}$$

One can see in (II.5) how (U_n) regulates the process (X_n). On the one hand, (U_n) generates offsprings if there is not enough "energy" in the process. On the other hand, (U_n) eliminates them if there is too much "energy" in the process. Via this adaptive control, the system is always "persistently excited," i.e., for all $n \in \mathbb{N}$, $a_n \geq 1$. The performance of the tracking can be evaluated by the average weighted cost sequence (C_n) given by

$$C_n = \frac{1}{n} \sum_{k=1}^{n} a_k^{-1}(X_k - x_k)^2.$$

The adaptive tracking is said to be optimal if C_n converges a.s. to σ^2, whereas it is globally stable if $\lim\sup C_n < \infty$ a.s. The following asymptotic properties for $(\hat{m}_n)$ were recently established in [6], while those concerning $(\hat{\sigma}_n^2)$ are new.

Theorem 1: Assume that $(Y_{n,i})$ has a finite moment of order > 2 and that (x_n) converges a.s. to an integer $x \geq 0$. If we use the adaptive control given by (II.5), then $\hat{m}_n$ is a strongly consistent estimator of m

$$(\hat{m}_n - m)^2 = O\left(\frac{\log n}{n}\right) \quad \text{a.s.} \tag{II.6}$$

In addition, if $\alpha = \max(1, P(m^{-1}x))$, we have the CLT

$$\sqrt{n}(\hat{m}_n - m) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2}{\alpha}\right) \tag{II.7}$$

and the LIL

$$\limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} (\hat{m}_n - m) = -\liminf_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} (\hat{m}_n - m) = \frac{\sigma}{\sqrt{\alpha}} \quad \text{a.s.} \tag{II.8}$$

In particular,

$$\limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right) (\hat{m}_n - m)^2 = \frac{\sigma^2}{\alpha} \quad \text{a.s.} \tag{II.9}$$

Finally, we have the quadratic strong law

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} (\hat{m}_k - m)^2 = \frac{\sigma^2}{\alpha} \quad \text{a.s.} \tag{II.10}$$

Theorem 2: Assume that $(Y_{n,i})$ has a finite moment of order > 2 and that (x_n) converges a.s. to an integer $x \geq 0$. If we use the adaptive control given by (II.5), then $\hat{\sigma}_n^2$ is a strongly consistent estimator of σ^2

$$|\hat{\sigma}_n^2 - \sigma_n^2| = O\left(\frac{\log n}{n}\right) \quad \text{a.s.} \tag{II.11}$$

In addition, assume that $(Y_{n,i})$ has a finite moment of order > 4. Denote by τ^4 the fourth order centered moment of $(Y_{n,i})$, and set $\rho = \alpha^{-1}\tau^4 + (2 - 3\alpha^{-1})\sigma^4$. Then, we have the CLT

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \rho) \tag{II.12}$$

and the LIL

$$\limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} (\hat{\sigma}_n^2 - \sigma^2) = -\liminf_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} (\hat{\sigma}_n^2 - \sigma^2) = \sqrt{\rho} \quad \text{a.s.} \tag{II.13}$$

In particular,

$$\limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right) (\hat{\sigma}_n^2 - \sigma^2)^2 = \rho \quad \text{a.s.} \tag{II.14}$$

Finally, we have the quadratic strong law

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} (\hat{\sigma}_k^2 - \sigma^2)^2 = \rho \quad \text{a.s.} \tag{II.15}$$

Remark 1: If $(Y_{n,i})$ has only a finite moment of order 2, then (II.6) and (II.11) hold replacing the convergence rates $O(\log n/n)$ by $o((\log n)^{1+\gamma}/n)$ for all $\gamma > 0$. Moreover, concerning the estimation of σ^2, similar results than those obtained for $\hat{\sigma}_n^2$ were proven in [6] for

$$\Gamma_n = \frac{1}{n} \sum_{k=1}^{n} a_k^{-1}(X_k - \hat{m}_{k-1}a_k)^2$$

and

$$\Sigma_n = \frac{1}{n} \sum_{k=1}^{n} a_k^{-1}(X_k - \hat{m}_k a_k)^2.$$

Finally, the tracking is globally stable and residually optimal since (C_n) converges a.s. to $\sigma^2 + \xi$, where $\xi = \alpha^{-1}(m\alpha - x)^2$, which differs from zero except for $x \in m\mathbb{N}^*$.

Proof: Theorem 1 is already established in [6] so that we only have to prove *Theorem 2*. For all $k \geq 1$, we have from (II.2), $X_k - \hat{m}_n a_k = (m - \hat{m}_n)a_k + \epsilon_k$. Consequently,

$$n(\hat{\sigma}_n^2 - \sigma_n^2) = (m - \hat{m}_n)^2 A_n + 2(m - \hat{m}_n) \sum_{k=1}^{n} \epsilon_k.$$

Moreover, it follows from (II.3) that $\sum_{k=1}^{n} \epsilon_k = A_n(m - \hat{m}_n)$. Therefore,

$$\sigma_n^2 - \hat{\sigma}_n^2 = \frac{A_n}{n}(\hat{m}_n - m)^2. \tag{II.16}$$

It was shown in [6] that both (a_n) and (A_n/n) converge a.s. to $\alpha = \max(1, P(m^{-1}x))$. Then, we directly deduce (II.11) from (II.16) and (II.6). Furthermore, it follows from the martingale CLT (see, e.g.,[17, Cor. 3.1]) that

$$\sqrt{n}(\sigma_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \rho) \tag{II.17}$$

where $\rho = \alpha^{-1}\tau^4 + (2 - 3\alpha^{-1})\sigma^4$. Moreover, we also obtain from the martingale LIL (see [26, Th. 3]) that

$$\limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} (\sigma_n^2 - \sigma^2) = -\liminf_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} (\sigma_n^2 - \sigma^2) = \sqrt{\rho} \quad \text{a.s.} \tag{II.18}$$

Then, (II.11) and (II.17) imply (II.12), whereas (II.13) is given by (II.11) and (II.18). Finally, it was shown in [6] that

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} (\sigma_k^2 - \sigma^2)^2 = \rho \quad \text{a.s.} \tag{II.19}$$

which directly implies (II.15) completing the *Proof* of *Theorem 2*. □

III. BGWI Results

We shall now focus our attention on the more attractive BGWI process

$$X_{n+1} = \sum_{i=1}^{X_n + U_n} Y_{n,i} + I_{n+1}. \tag{III.1}$$

As in the previous section, we assume that (U_n) is a sequence of integer-valued random variables adapted to $\mathbb{F}$ such that, whatever the value of $n \in \mathbb{N}$, $X_n + U_n \geq 1$. We can rewrite (III.1) as the autoregressive form

$$X_n = ma_n + \lambda + \epsilon_n, \quad a_n = X_{n-1} + U_{n-1} \tag{III.2}$$

where $\epsilon_n = X_n - ma_n - \lambda$. Then, (ϵ_n) is a martingale difference sequence adapted to $\mathbb{F}$ such that

$$E[\epsilon_n^2 | \mathcal{F}_{n-1}] = \sigma^2 a_n + b^2$$
$$E[a_n^{-1}\epsilon_n^2 | \mathcal{F}_{n-1}] \leq \sigma^2 + b^2 \tag{III.3}$$

as $a_n \geq 1$. Therefore, we have the stochastic regression equation

$$X_n = \theta^t \Phi_n + \epsilon_n \tag{III.4}$$

where $\theta^t = (m, \lambda)$ and $\Phi_n^t = (a_n, 1)$. In order to estimate the vector of means θ, we propose to make use of the WLS estimator that minimizes the quadratic criterion

$$\Delta_n(\theta) = \frac{1}{2} \sum_{k=1}^{n} \alpha_k (X_k - \theta^t \Phi_k)^2.$$

The choice of the weighted sequence (α_n) is crucial. We propose to take

$$\alpha_n^{-1} = a_n(\log A_n)^{1+\gamma} \quad \text{with} \quad A_n = \sum_{k=1}^{n} a_k + \gamma \tag{III.5}$$

where $\gamma > 0$. Then, we clearly have

$$\hat{\theta}_n = S_n^{-1} \sum_{k=1}^{n} \alpha_k \Phi_k X_k$$
$$S_n = \sum_{k=1}^{n} \alpha_k \Phi_k \Phi_k^t + S \tag{III.6}$$

where S is a deterministic, symmetric and positive-definite matrix. It is added to the matrix S_n in order to avoid useless invertibility assumptions. Similar WLS algorithms were first introduced by Bercu and Duflo [3]. In the ARMAX framework, it was shown [4], [5] that the WLS performs as well as the extended least squares (ELS) for parameter estimation. In addition, the WLS behaves better than the ELS for the tracking optimality. Finally, Guo [15] has recently proved the almost sure self-convergence of the WLS algorithm. This property is a key point of our approach. In order to estimate the vector of variances $\eta^t = (\sigma^2, b^2)$, we use the same ideas developed by Winnicki [31] (see also [20], [32]). First, we assume that θ is known. Then, we can set

$$\epsilon_n^2 = \eta^t \Phi_n + v_n \tag{III.7}$$

and we can study (III.7) as a stochastic regression equation with unknown parameter η. (v_n) is clearly a martingale difference sequence. Moreover, if τ^4 and ν^4 are the fourth order centered moments of $(Y_{n,i})$ and (I_n) respectively, we find that

$$E[v_n^2 | \mathcal{F}_{n-1}] = 2a_n^2\sigma^4 + a_n(\tau^4 - 3\sigma^4 + 4b^2\sigma^2) + \nu^4 - b^4.$$

Hence, as $a_n \geq 1$, we deduce that

$$E[a_n^{-2}v_n^2 | \mathcal{F}_{n-1}] \leq \tau^4 - \sigma^4 + 4b^2\sigma^2 + \nu^4 - b^4.$$

Therefore, we are led to introduce the weighted sequence (β_n) such that $\beta_n = a_n^{-1}\alpha_n$ and we propose to make use of the WLS estimator

$$\eta_n = Q_n^{-1} \sum_{k=1}^{n} \beta_k \Phi_k \epsilon_k^2$$
$$Q_n = \sum_{k=1}^{n} \beta_k \Phi_k \Phi_k^t + Q \qquad \text{(III.8)}$$

where Q is a deterministic, symmetric and positive definite matrix. Finally, as the vector θ is unknown, we propose to choose

$$\hat{\eta}_n = Q_n^{-1} \sum_{k=1}^{n} \beta_k \Phi_k \hat{\epsilon}_k^2 \qquad \text{(III.9)}$$

where $\hat{\epsilon}_k = X_k - \hat{\theta}_n^t \Phi_k$. Furthermore, proceeding as in Section II, we propose to make use of the adaptive tracking control

$$U_n = \begin{cases} 1 - X_n, & \text{if } P(\hat{m}_n^{-1}(x_{n+1} - \hat{\lambda}_n)) = 0 \\ P(\hat{m}_n^{-1}(x_{n+1} - \hat{\lambda}_n)) - X_n, & \text{otherwise} \end{cases} \qquad \text{(III.10)}$$

where $\hat{\theta}_n^t = (\hat{m}_n, \hat{\lambda}_n)$.

Lemma 1: Assume that (x_n) converges a.s. to an integer $x \geq 0$. If we use the adaptive control given by (III.10), then there exists a finite random variable $l \geq 1$ such that a.s.

$$(\log n)^{1+\gamma} \frac{S_n}{n} \longrightarrow L = \begin{bmatrix} l & 1 \\ 1 & l^{-1} \end{bmatrix}. \qquad \text{(III.11)}$$

Remark 2: First, we want to point out that we have only required moments of order two for $(Y_{n,i})$ and (I_n). Next, as $\det(L) = 0$, $\hat{\theta}_n$ is not a consistent estimator of θ. In fact, it is possible to show the sharper result

$$\left\| (\log n)^{1+\gamma} \frac{S_n}{n} - L \right\| = O\left(\frac{1}{\log n}\right) \quad \text{a.s.}$$

Therefore,

$$\|L^{1/2}(\hat{\theta}_n - \theta)\|^2 = O\left(\frac{1}{\log n}\right) \quad \text{a.s.}$$

which implies that $(l\hat{m}_n + \hat{\lambda}_n)$ converges to $lm + \lambda$ a.s.

Remark 3: Assume that both $(Y_{n,i})$ and (I_n) possess finite moments of order > 2. Then, we have, via Chow's Lemma (see e.g. [12, Prop. 1.3.19]), that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} a_k^{-1} \epsilon_k^2 \leq \sigma^2 + b^2 \quad \text{a.s.}$$

Moreover, it is shown in the *Proof* of *Lemma 1* that (a_n) converges a.s. to l. Hence, (III.2) immediately implies that

$$\limsup_{n\to\infty} C_n \leq \sigma^2 + b^2 + l^{-1}(ml + \lambda - x)^2 \quad \text{a.s.}$$

and the tracking is globally stable.

In order to obtain the strong consistency for $(\hat{\theta}_n)$, we are led to introduce an exogenous excitation on the adaptive tracking control. This approach is frequently used in adaptive tracking whenever there is a lack of energy which does not allow us to properly estimate the parameters (see e.g., [4], [5], [13], [14], [22], and [23]). The effect of this excitation will be to make the limit matrix L in (III.11) invertible. Consequently, we propose to make use of the continually disturbed adaptive tracking control

$$U_n = P(\hat{m}_n^{-1}(x_{n+1} - \hat{\lambda}_n)) - X_n + V_n \qquad \text{(III.12)}$$

where (V_n) is an exogenous bounded sequence of i.i.d. positive integer-valued random variables, adapted to $\mathbb{F}$. In addition, we assume that (V_n) is independent of $(Y_{n,i})$, (I_n), (x_n) and of the initial variables X_0 and U_0. We denote by V the nondegenerate distribution of (V_n).

Lemma 2: Assume that (x_n) converges a.s. to an integer $x \geq 0$. If we use the adaptive control given by (III.12), then, for $h = P(m^{-1}(x - \lambda))$, we have a.s.

$$(\log n)^{1+\gamma} \frac{S_n}{n} \longrightarrow H = \begin{bmatrix} E[h+V] & 1 \\ 1 & E[(h+V)^{-1}] \end{bmatrix}. \qquad \text{(III.13)}$$

Remark 4: The matrix H is invertible since, by Jensen's inequality and nondegeneracy of V, $\det(H) = E[h+V]E[(h+V)^{-1}] - 1 > 0$. Moreover, it is shown in the *Proof* of *Lemma 2* that (A_n/n) converges a.s. to $h + E[V]$. Hence, as before, (III.2) immediately implies that the tracking is globally stable.

Theorem 3: Assume that (x_n) converges a.s. to an integer $x \geq 0$. If we use the adaptive control given by (III.12), then $\hat{\theta}_n$ is a strongly consistent estimator of θ

$$\|\hat{\theta}_n - \theta\|^2 = O\left(\frac{(\log n)^{1+\gamma}}{n}\right) \quad \text{a.s.} \qquad \text{(III.14)}$$

In addition, assume that both $(Y_{n,i})$ and (I_n) possess finite moments of order > 2. Then, we have the CLT

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, H^{-1}WH^{-1}) \qquad \text{(III.15)}$$

where

$$W = \begin{bmatrix} E[\sigma^2(h+V) + b^2] & E\left[\dfrac{\sigma^2(h+V)+b^2}{(h+V)}\right] \\ E\left[\dfrac{\sigma^2(h+V)+b^2}{(h+V)}\right] & E\left[\dfrac{\sigma^2(h+V)+b^2}{(h+V)^2}\right] \end{bmatrix}.$$

Moreover, for any vector $u \in \mathbb{R}^2$, we have the LIL

$$\limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} u^t(\hat{\theta}_n - \theta)$$
$$= -\liminf_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} u^t(\hat{\theta}_n - \theta)$$
$$= (u^t H^{-1} W H^{-1} u)^{1/2} \quad \text{a.s.} \qquad \text{(III.16)}$$

In particular, if $\lambda_{\min}(H^{-1}WH^{-1})$ and $\lambda_{\max}(H^{-1}WH^{-1})$ are the minimum and the maximum eigenvalues of $H^{-1}WH^{-1}$ respectively, then we have

$$\lambda_{\min}(H^{-1}WH^{-1}) \leq \limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right) \|\hat{\theta}_n - \theta\|^2 \leq \lambda_{\max}(H^{-1}WH^{-1}) \quad \text{a.s.}$$

Lemma 3: Assume that (x_n) converges a.s. to an integer $x \geq 0$ and that both $(Y_{n,i})$ and (I_n) possess finite moments of order

4. If we use the adaptive control given by (III.12), then we have a.s.

$$(\log n)^{1+\gamma}\frac{Q_n}{n} \longrightarrow K = \begin{bmatrix} 1 & E[(h+V)^{-1}] \\ E[(h+V)^{-1}] & E[(h+V)^{-2}] \end{bmatrix}. \tag{III.17}$$

Remark 5: The matrix K is also invertible since by the Cauchy–Schwarz inequality and the nondegeneracy of V, $\det(K) = E[(h+V)^{-2}] - E^2[(h+V)^{-1}] > 0$.

Theorem 4: Assume that (x_n) converges a.s. to an integer $x \geq 0$ and that both $(Y_{n,i})$ and (I_n) possess finite moments of order 4. If we use the adaptive control given by (III.12), then $\hat{\eta}_n$ is a strongly consistent estimator of η

$$\|\hat{\eta}_n - \eta\|^2 = O\left(\frac{(\log n)^{1+\gamma}}{n}\right) \quad \text{a.s.} \tag{III.18}$$

In addition, assume that both $(Y_{n,i})$ and (I_n) possess finite moments of order > 4. Then, we have the CLT

$$\sqrt{n}(\hat{\eta}_n - \eta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, K^{-1}PK^{-1}) \tag{III.19}$$

where

$$P = \begin{bmatrix} E\left[\frac{R(h+V)}{(h+V)^2}\right] & E\left[\frac{R(h+V)}{(h+V)^3}\right] \\ E\left[\frac{R(h+V)}{(h+V)^3}\right] & E\left[\frac{R(h+V)}{(h+V)^4}\right] \end{bmatrix}$$

and $R(z) = 2z^2\sigma^4 + z(\tau^4 - 3\sigma^4 + 4b^2\sigma^2) + \nu^4 - b^4$. Moreover, for any vector $u \in \mathbb{R}^2$, we have the LIL

$$\begin{aligned} \limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} u^t(\hat{\eta}_n - \eta) &= -\liminf_{n\to\infty} \left(\frac{n}{2\log\log n}\right)^{1/2} u^t(\hat{\eta}_n - \eta) \\ &= (u^t K^{-1}PK^{-1}u)^{1/2} \quad \text{a.s.} \end{aligned} \tag{III.20}$$

In particular,

$$\begin{aligned} \lambda_{\min}(K^{-1}PK^{-1}) &\leq \limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right) \|\hat{\eta}_n - \eta\|^2 \\ &\leq \lambda_{\max}(K^{-1}PK^{-1}) \quad \text{a.s.} \end{aligned}$$

IV. Conclusion

In this paper, we expanded the adaptive control theory to the branching processes framework. For branching processes with immigration, we have shown that, thanks to the introduction of a suitable adaptive control, it was possible to consistently estimate all the parameters of interest m, λ, σ^2, and b^2 without any restriction on the range of m. BGWI processes appear in a wide range of applications going from population biology and statistical physics to traffic flow and computer sciences. We refer the reader to Heyde and Seneta [18] for a short historical discussion and to [10], [16], and [30] for more recent contributions. It is reasonable to think that some practical applications of our approach are possible, such as the following:

- in statistical physics with the control of the total number of particles contained in a small volume;
- in computer sciences with the electronic management of documents in Web environment.

The investigation of one of these two applications can be the purpose of a new development in the area of tracking control for BGWI processes.

Appendix A
Weighted Law of Large Numbers

We shall often make use of the weighted law of large numbers (WLLN) for martingales established by Duflo and Bercu [3], [12]. We also mention the important self-convergence property associated with this weighted law, which was recently proven by Guo [15]. Let (ξ_n) be a sequence of random variables, adapted to $\mathbb{F}$, such that a.s.

$$E[\xi_n|\mathcal{F}_{n-1}] = 0$$

and

$$\sup_n E[\xi_n^2|\mathcal{F}_{n-1}] < \infty.$$

Let (ϕ_n) be a sequence of random vectors of dimension $d \geq 1$ such that ϕ_n is $\mathcal{F}_{n-1}$ measurable. For $n \geq 0$, set

$$\begin{aligned} M_n &= M_0 + \sum_{k=1}^{n} \gamma_k \phi_k \xi_k \\ S_n &= S_0 + \sum_{k=1}^{n} \gamma_k \phi_k \phi_k^t \end{aligned} \tag{A.1}$$

where S_0 is a deterministic, symmetric and positive definite matrix. Denote by $\lambda_{\min} S_n$ and $\lambda_{\max} S_n$ the minimum and the maximum eigenvalues of S_n, respectively.

Theorem 5: Assume that (γ_n) is an admissible weighted sequence, i.e. (γ_n) is a decreasing bounded sequence such that γ_n is $\mathcal{F}_{n-1}$ measurable and almost surely

$$\sum_{n=1}^{\infty} \gamma_n f_n < \infty \quad \text{with} \quad f_n = \gamma_n \phi_n^t S_n^{-1} \phi_n. \tag{A.2}$$

Then, if $Z_n = S_n^{-1} M_n$, the sequence (Z_n) converges a.s. to a finite random vector Z. In addition, on the set

$$I = \{\lim_{n\to\infty} \lambda_{\min} S_n = +\infty\}$$

(Z_n) converges a.s. to zero with the almost sure rate of convergence

$$\|Z_n\|^2 = O\left(\frac{1}{\lambda_{\min} S_n}\right) \quad \text{a.s.} \tag{A.3}$$

Remark 6: We have the very appropriate identity

$$f_n = \frac{\det S_n - \det S_{n-1}}{\det S_n}.$$

Consequently, if $(\log\det S_n)^{1+\gamma} = O(\gamma_n^{-1})$ with $\gamma > 0$, then (A.2) is always satisfied as

$$\sum_{n=1}^{\infty} \gamma_n f_n = O\left(\sum_{n=1}^{\infty} \frac{\det S_n - \det S_{n-1}}{\det S_n (\log\det S_n)^{1+\gamma}}\right) < \infty.$$

Appendix B
Proof of Lemmas 1 and 2

From relations (III.4) and (III.6)

$$\hat{\theta}_n - \theta = S_n^{-1} M_n, \quad \text{where}$$
$$M_n = M_0 + \sum_{k=1}^{n} \alpha_k \Phi_k \epsilon_k \tag{B.1}$$

and $M_0 = -S\theta$. We want to apply the WLLN to $\gamma_n^{-1} = (\log A_n)^{1+\gamma}$

$$\phi_n = \frac{\Phi_n}{\sqrt{a_n}} \quad \text{and} \quad \xi_n = \frac{\epsilon_n}{\sqrt{a_n}}.$$

First, we have already seen that $E[\xi_n^2|\mathcal{F}_{n-1}] \le \sigma^2 + b^2$. In addition, we have to check that (A.2) is satisfied. In fact, it is easy to see that $\det S_n \le (\lambda_{\max} S_n)^2$ and $\lambda_{\max} S_n \le 2A_n$ so that $\log\det S_n \le 2\log(2A_n)$ which clearly implies (A.2) via Remark 6. It follows from the first part of *Theorem 5* that there exists a finite random vector $\overline{\theta}$ such that $\hat{\theta}_n$ converges a.s. to $\overline{\theta}$. Set $l_n = \max(1, h_n)$ where $h_n = P(\hat{m}_n^{-1}(x_{n+1} - \hat{\lambda}_n))$. On the one hand, for the adaptive tracking control (III.10) situation without excitation, the sequences (a_n) and (A_n/n) both converge a.s. to the finite random variable $l = \max(1, h)$ where $h = P(\overline{m}^{-1}(x - \overline{\lambda}))$ and $\overline{\theta}^t = (\overline{m}, \overline{\lambda})$. The rest of the proof follows essentially the same lines as in [5, App. B]. We can use the Abel relation

$$S_n = \gamma_{n+1} R_n + \sum_{k=1}^{n} \delta_k \frac{R_k}{k} + S \tag{B.2}$$

with

$$R_n = \sum_{k=1}^{n} a_k^{-1} \Phi_k \Phi_k^t, \quad \delta_n = n(\gamma_n - \gamma_{n+1}). \tag{B.3}$$

We have already shown that a.s.

$$\frac{1}{n} R_n \longrightarrow L = \begin{bmatrix} l & 1 \\ 1 & l^{-1} \end{bmatrix}. \tag{B.4}$$

In addition, as γ_n^{-1} is almost surely equivalent to $(\log n)^{1+\gamma}$

$$\sum_{k=1}^{n} \delta_k \sim (1+\gamma) \frac{n\gamma_n}{\log n} \quad \sum_{k=1}^{n} \delta_k = o(n\gamma_n) \quad \text{a.s.}$$

Therefore, by use of (B.2), (B.4), and Toeplitz's Lemma (see, e.g., [12, p. 54]), we obtain the convergence (III.11) of *Lemma 1*. On the other hand, for the adaptive tracking control (III.12) situation with excitation, we have already seen that h_n converges a.s. to a finite random variable h. First, assume that h is a deterministic constant. Then, (A_n/n) converges a.s. to $h + E[V]$ and we also have a.s.

$$\frac{1}{n} R_n \longrightarrow H = \begin{bmatrix} E[h+V] & 1 \\ 1 & E[(h+V)^{-1}] \end{bmatrix}. \tag{B.5}$$

By Jensen's inequality, the matrix H is invertible so that $n = O(\lambda_{\min} R_n)$ a.s. Next, if h is a random variable, h has necessarily a discrete sample space. In addition, for all n greater than a finite random integer n_0, $h_n = h$ and the exogenous excitation $(V_n)_{n \ge n_0}$ is independent of h. Thus, as Jensen's inequality holds for conditional expectation, we also obtain via a partition argument that $n = O(\lambda_{\min} R_n)$ a.s. Therefore, as the weighted sequence (γ_n) is decreasing, it results that $\gamma_n R_n \le S_n$ so that $n\gamma_n = O(\lambda_{\min} S_n)$ a.s. Consequently, as γ_n^{-1} is almost surely equivalent to $(\log n)^{1+\gamma}$, we find that $(\lambda_{\min} S_n)$ goes a.s. to infinity and we can deduce from *Theorem 5* that $\hat{\theta}_n$ is a strongly consistent estimator of θ. Finally, the value of h is the deterministic constant $h = P(m^{-1}(x - \lambda))$ which achieves the proof of the convergence (III.13) of *Lemma 2*. □

Appendix C
Proof of Theorem 3

We can immediately deduce (III.14) from relation (A.3) as

$$||\hat{\theta}_n - \theta||^2 = O\left(\frac{1}{\lambda_{\min} S_n}\right) \quad \text{a.s.} \tag{C.1}$$

and $n\gamma_n = O(\lambda_{\min} S_n)$ a.s. For the CLT, we use the same approach as Klimko and Nelson [21] or Wei and Winnicki [29] for BGWI processes without control. From (B.1), we have $\sqrt{n}(\hat{\theta}_n - \theta) = H_n^{-1} Z_n$ where

$$H_n = \frac{1}{n\gamma_n} S_n \quad \text{and} \quad Z_n = \frac{1}{\sqrt{n\gamma_n}} M_n.$$

In order to establish (III.15), we only have to prove that

$$Z_n \xrightarrow{\mathcal{L}} N(0, W)$$

since H_n converges to H a.s. By the Cramer-Wold Theorem (see, e.g., [17, p. 175]), we have to show that for any $u \in \mathbb{R}^2$ with $u \neq 0$,

$$\langle u, Z_n \rangle \xrightarrow{\mathcal{L}} N(0, u^t W u). \tag{C.2}$$

By (III.3), $\langle u, M_n \rangle$ is a martingale with increasing process $u^t W_n u$ such that

$$W_n = \sum_{k=1}^{n} a_k^{-2} \gamma_k^2 (\sigma^2 a_k + b^2) \Phi_k \Phi_k^t.$$

Exactly as in Appendix B, we can prove that a.s.

$$\frac{1}{n\gamma_n^2} W_n \longrightarrow W = \begin{bmatrix} E[\sigma^2(h+V) + b^2] & E\left[\frac{\sigma^2(h+V) + b^2}{(h+V)}\right] \\ E\left[\frac{\sigma^2(h+V) + b^2}{(h+V)}\right] & E\left[\frac{\sigma^2(h+V) + b^2}{(h+V)^2}\right] \end{bmatrix}. \tag{C.3}$$

Moreover, assume that $(Y_{n,i})$ and (I_n) both possess finite moments of order $2+2\delta$ with $\delta > 0$. From (III.2) and the Rosenthal inequality (see e.g. [25, Th. 2.12]), we have for all $n \geq 1$, $E[|\epsilon_n|^{2+2\delta}|\mathcal{F}_{n-1}] = O(a_n^{1+\delta})$. Therefore, we have a.s.

$$\begin{aligned}\sum_{k=1}^{n} E\left[\frac{\gamma_k}{a_k}\langle u, \Phi_k\rangle\epsilon_k|^{2+2\delta}|\mathcal{F}_{k-1}\right] &= O\left(\sum_{k=1}^{n}\left(\frac{\gamma_k}{a_k}||\Phi_k||\right)^{2+2\delta} a_k^{1+\delta}\right)\\ &= O(n\gamma_n^{2+2\delta}) = o(n\gamma_n^2). \quad \text{(C.4)}\end{aligned}$$

Relations (C.3) and (C.4) show that the conditions of the martingale CLT are satisfied (see e.g. [17, Cor. 3.1]) which achieves the proof of (C.2). Finally, via the martingale LIL (see e.g. [26, Th. 3]), we have for any $u \in \mathbb{R}^2$ with $u \neq 0$

$$\begin{aligned}&\limsup_{n\to\infty}\left(\frac{1}{2n\gamma_n^2\log\log n}\right)^{1/2}\langle u, M_n\rangle\\ &= -\liminf_{n\to\infty}\left(\frac{1}{2n\gamma_n^2\log\log n}\right)^{1/2}\langle u, M_n\rangle\\ &= (u^tWu)^{1/2} \quad \text{a.s.} \quad \text{(C.5)}\end{aligned}$$

since u^tW_nu is a.s. equivalent to $n\gamma_n^2u^tWu$. As $M_n = S_n(\hat{\theta}_n - \theta)$, (III.16) follows from *Lemma 2* together with (C.5) which completes the *Proof* of *Theorem 3*. □

APPENDIX D
PROOF OF THEOREM 4

From relations (III.7) and (III.8), we have

$$\eta_n - \eta = Q_n^{-1}M_n \quad \text{where} \quad M_n = M_0 + \sum_{k=1}^{n}\beta_k\Phi_kv_k \quad \text{(D.1)}$$

and $M_0 = -Q\eta$. It is possible to apply the WLLN to $\gamma_n^{-1} = (\log A_n)^{1+\gamma}$

$$\phi_n = \frac{\Phi_n}{a_n} \quad \text{and} \quad \xi_n = \frac{v_n}{a_n}.$$

First, we have already seen that $E[\xi_n^2|\mathcal{F}_{n-1}] \leq \tau^4 + 4b^2\sigma^2 + \nu^4$. In addition, since $\log\det Q_n \leq 2\log(2A_n)$, condition (A.2) is satisfied. Therefore, we prove the convergence (III.17) of *Lemma 3* following exactly the same lines as in Appendix B. Thus, we deduce from (A.3) that

$$||\eta_n - \eta||^2 = O\left(\frac{(\log n)^{1+\gamma}}{n}\right) \quad \text{a.s.} \quad \text{(D.2)}$$

On the other hand, we have from (III.2), (III.8), and (III.9)

$$\begin{aligned}Q_n(\hat{\eta}_n - \eta_n) &= \sum_{k=1}^{n}\beta_k\Phi_k(\hat{\epsilon}_k^2 - \epsilon_k^2),\\ &= \sum_{k=1}^{n}\beta_k\Phi_k[(\hat{\theta}_n - \theta)^t\Phi_k]^2\\ &\quad - 2\sum_{k=1}^{n}\beta_k\Phi_k[(\hat{\theta}_n - \theta)^t\Phi_k]\epsilon_k. \quad \text{(D.3)}\end{aligned}$$

Consequently, using again the WLLN for the martingale in (D.3), we find that

$$\begin{aligned}&||Q_n(\hat{\eta}_n - \eta_n)||\\ &\quad = O(n\gamma_n||\hat{\theta}_n - \theta||^2) + O(\sqrt{n\gamma_n||\hat{\theta}_n - \theta||^2}) \quad \text{a.s.}\end{aligned}$$

Hence, we obtain from the LIL (III.16) together with (III.17) that

$$||\hat{\eta}_n - \eta_n||^2 = o\left(\frac{(\log n)^{2+2\gamma}}{n^2}\right) \quad \text{a.s.} \quad \text{(D.4)}$$

Therefore, (III.18) is directly given by (D.2) and (D.4). For the CLT, using the same approach as in Appendix C, we find that

$$\sqrt{n}(\eta_n - \eta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, K^{-1}PK^{-1}). \quad \text{(D.5)}$$

Then, (III.19) immediately follows from (D.4) and (D.5). Finally, for any $u \in \mathbb{R}^2$ with $u \neq 0$, we also have

$$\begin{aligned}&\limsup_{n\to\infty}\left(\frac{1}{2n\gamma_n^2\log\log n}\right)^{1/2}\langle u, M_n\rangle\\ &= -\liminf_{n\to\infty}\left(\frac{1}{2n\gamma_n^2\log\log n}\right)^{1/2}\langle u, M_n\rangle\\ &= (u^tPu)^{1/2} \quad \text{a.s.} \quad \text{(D.6)}\end{aligned}$$

with $M_n = Q_n(\eta_n - \eta)$. Consequently, we obtain (III.20) via *Lemma 3* together with (D.4) and (D.6) which completes the *Proof* of *Theorem 4*. □

ACKNOWLEDGMENT

The author wishes to express his gratitude to B. Duncan, D. Duncan, and T. Duncan for their very warm hospitality and many helpful discussions while he was visiting the University of Kansas.

REFERENCES

[1] S. Asmussen and H. Hering, *Branching Processes*. Boston, MA: Birkhäuser, 1983.
[2] K. B. Athreya and P. E. Ney, *Branching Processes*. Berlin, Germany: Springer-Verlag, 1972.
[3] B. Bercu and M. Duflo, "Moindres carrés pondérés eet poursuite," *Ann. Inst. Henri Poincaré*, vol. 28, pp. 403–430, 1992.
[4] B. Bercu, "Weighted estimation and tracking for ARMAX models," *SIAM J. Control Optim.*, vol. 33, pp. 89–106, 1995.
[5] ——, "Central limit theorem and law of iterated logarithm for the least squares algorithms in adaptive tracking," *SIAM J. Control Optim.*, vol. 36, pp. 910–928, 1998.
[6] ——, "Weighted estimation and tracking for Bienaymé Galton Watson processes with adaptive control," *Statist. Probab. Lett.*, vol. 42, pp. 415–421, 1999.
[7] P. E. Caines, *Linear Stochastic Systems*. New York: Wiley, 1988.
[8] H. F. Chen and L. Guo, *Identification and Stochastic Adaptive Control*. Boston, MA: Birkhäuser, 1991.
[9] Y. S. Chow, "Local convergence of martingale and the law of large numbers," *Ann. Math. Statist.*, vol. 36, pp. 552–558, 1965.
[10] D. A. Dawson and Y. Wu, "Multilevel multitype branching models of an information system," in *IMA Vol. Math. Appl*. Berlin, Germany: Springer-Verlag, 1997, vol. 84, pp. 57–72.
[11] J. P. Dion and N. M. Yanev, "Limit theorems and estimation theory for branching processes with and increasing random number of ancestors," *J. Appl. Probab.*, vol. 34, pp. 309–327, 1997.
[12] M. Duflo, *Random Iterative Models*. Berlin, Germany: Springer-Verlag, 1997.

[13] L. Guo and H. F. Chen, “The Aström–Wittenmark self-tuning regulator revisited and ELS-based adaptive trackers,” *IEEE Trans. Automat. Contr.*, vol. 36, pp. 802–812, 1991.
[14] L. Guo, “Further results on least squares based adaptive minimum variance control,” *SIAM J. Control Optim.*, vol. 32, pp. 187–212, 1994.
[15] ——, “Self convergence of weighted least squares with applications to stochastic adaptive control,” *IEEE Trans. Automat. Contr.*, vol. 41, pp. 79–89, 1996.
[16] P. Guttorp, *Statistical Inference for Branching Processes*. New York: Wiley, 1991.
[17] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application*. New York: Academic, 1980.
[18] C. C. Heyde and E. Seneta, “Estimation theory for growth and immigration rates in a multiplicative process,” *J. Appl. Probab.*, vol. 9, pp. 235–258, 1972.
[19] C. C. Heyde and E. Seneta, “Notes on "Estimation theory for growth and immigration rates in a multiplicative process,” *J. Appl. Probab.*, vol. 11, pp. 572–577, 1974.
[20] C. C. Heyde, “On estimating the variance of the offspring distribution in a simple branching process,” *Adv. Appl. Probab.*, vol. 6, pp. 421–433, 1974.
[21] L. A. Klimko and P. I. Nelson, “On conditional least squares estimation for stochastic processes,” *Ann. Statist.*, vol. 6, pp. 629–642, 1978.
[22] T. L. Lai and C. Z. Wei, “Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems,” *Ann. Statist.*, vol. 10, pp. 154–166, 1982.
[23] ——, “On the concept of excitation in least squares identification and adaptive control,” *Stochastics*, vol. 16, pp. 227–254, 1986.
[24] A. G. Pakes, “On the critical Galton-Watson process with immigration,” *J. Austral. Math. Soc.*, vol. 12, pp. 476–482, 1971.
[25] V. V. Petrov, *Limit Theorems of Probability Theory*. Oxford, U. K.: Oxford Univ. Press, 1995.
[26] W. F. Stout, “A martingale analogue of Kolmogorov’s law of the iterated logarithm,” *Z. Wahrscheinlichkeitstheorie*, vol. 15, pp. 279–290, 1970.
[27] C. Z. Wei, “Adaptive prediction by least squares predictors in stochastic regression models with applications to time series,” *Ann. Statist.*, vol. 15, pp. 1667–1682, 1987.
[28] C. Z. Wei and J. Winnicki, “Some asymptotic results for branching processes with immigration,” *Stochastic Process. Appl.*, vol. 31, pp. 261–282, 1989.
[29] ——, “Estimation of the means in the branching process with immigration,” *Ann. Statist.*, vol. 18, pp. 1757–1773, 1990.
[30] J. Winnicki, “Estimation theory for the branching process with immigration,” *Contemp. Math.*, vol. 80, pp. 301–322, 1988.
[31] ——, “Estimation of the variances in the branching process with immigration,” *Probab. Theory Rel. Fields*, vol. 88, pp. 77–106, 1991.
[32] N. M. Yanev and S. Tchoukova-Dantcheva, “On the statistics of branching processes with immigration,” *C.R. Acad. Bulg. Sci.*, vol. 33, pp. 463–471, 1980.

Bernard Bercu was born in Paris, France, in 1965. He received the Ph.D. degree in mathematics from the University of Paris-Sud, Orsay, France, in 1992.

He is currently Maître de Conférences at the University of Paris-Sud. His research interests include stochastic algorithms and martingales, estimation theory and adaptive tracking control, branching processes, and large deviations applied to statistics.