Electronic Journal of Statistics
Vol. 11 (2017) 295–325
ISSN: 1935-7524
DOI: 10.1214/17-EJS1232

Minimum disparity estimation in controlled branching processes

Miguel González[*], **Carmen Minuesa**[*] **and Inés del Puerto**[*]

Department of Mathematics and Instituto de Computación Científica Avanzada (ICCAEx), University of Extremadura, Avda. de Elvas s/n, 06006 Badajoz, Spain
e-mail: mvelasco@unex.es; cminuesaa@unex.es; idelpuerto@unex.es

Abstract: Minimum disparity estimation in controlled branching processes is dealt with by assuming that the offspring law belongs to a general parametric family. Under some regularity conditions it is proved that the minimum disparity estimators proposed -based on the nonparametric maximum likelihood estimator of the offspring law when the entire family tree is observed- are consistent and asymptotic normally distributed. Moreover, the robustness of the estimators proposed is discussed. Through a simulated example, focusing on the minimum Hellinger and negative exponential disparity estimators, it is shown that both are robust against outliers, and the minimum negative exponential estimator is also robust against inliers.

MSC 2010 subject classifications: 60J80, 62M05.
Keywords and phrases: Branching process, controlled process, minimum disparity estimation, robustness.

Received January 2016.

1. Introduction

Branching processes are useful models for the description of the dynamics of systems whose elements produce new ones following probability laws. Its theory has been developed from simple models to increasing realism. Added to the theoretical interest in these processes there is therefore a major practical dimension due to their potential applications in such diverse fields as biology, epidemiology, genetics, medicine, nuclear physics, demography, actuarial mathematics, algorithm and data structures, see, for example, the monographs [9, 23, 25, 29, 34].

In particular, controlled branching processes (CBPs) are discrete time stochastic processes very appropriate to describe the growth of populations in which the number of participating individuals in the reproduction process is determined in each generation by a control mechanism. Besides, as is common in the branching framework, every individual reproduces independently of the others following the same probability law, which is called the offspring distribution. The novelty of adding to the branching notion a mechanism that fixes the number of progenitors generation by generation allows to model a great variety of random migratory movements (immigration, emigration, or even both

[*]Fax: +34 924272911, Tel.: +34 924289300; Ext.: 86115 (Miguel González); Ext.: 86108 (Carmen Minuesa); Ext.: 86820 (Inés del Puerto).

depending on the generation sizes). Applications can be found, for example, in an ecological context. Consider an invasive animal species that is widely recognized as a threat to native ecosystems, but there is disagreement about plans to eradicate it, i.e., while the presence of the species is appreciated by a part of the society, if its numbers are left uncontrolled it is known to be very harmful to native ecosystems. In such a case, it is better to control the population to keep it between admissible limits (a deterministic control function can be appropriated) even though this might mean periods when animals have to be culled. Two examples of discussions about this topic are [13, 44]. Another practical situation that can be modelled by this kind of process is the evolution of an animal population that is threatened by the existence of predators. In each generation, the survival of each animal (and therefore the possibility of giving new births) will be strongly affected by this factor, making the introduction of a random mechanism necessary to model the evolution of this kind of population. In such a situation a binomial control process would be reasonable or its approximation by Poisson distributions as the survival probability of an animal is very low and the reproductive capacity is high (see Section 2 for details).

Several well-known branching processes can be included in this class as particular cases by considering specific control mechanisms, for instance, the own Bienaymé–Galton–Watson (BGW) process, the branching processes with immigration (see [40] and references therein), with random migration (see [46]), with immigration at state zero (see [7]) or with bounded emigration (see [35]). Other interesting particular cases are branching processes with adaptive control (see [4]) or with continuous state space (see [36]).

Since the appearance of the pioneering publication by [47], the probability theory of the CBPs has been widely studied, showing increased diversity of behaviours far from those of the classical branching models, and, then, becoming these processes an interesting and flexible tool for describing more complex situations (see, for instance [21]). The development of its inference theory, which guarantees the applicability of these models, has become the main goal in the most recent researches. For this issue, in a frequentist framework, it is important to mention [12, 16, 17, 18, 28, 41]. From a Bayesian standpoint, one can find the papers [14, 15, 31].

From an applied standpoint, it is of interest to develop robust procedures in the field of branching processes due to the fact that it is not unusual to find in the development of a population the existence of a small proportion of individuals whose reproductive capacity is influenced by temporary events that can provoke outliers in the model. For instance, this may happen due to the presence of a disease with a low prevalence or punctual changes on the environmental conditions. These situations can hide the assumption of common offspring distribution along generations, usually assumed in the context of branching processes. Moreover, maximum likelihood estimation is badly affected by outliers as is well-known and pointed out in the simulated example at the end of the paper. Until now, robust estimation has barely developed in this context. One only can find results in the frame of the BGW processes, by using weighted least trimmed estimation (see [43]) and by considering minimum

Hellinger distance estimation (see [42]). The aforementioned facts motivate the need to go in depth in robust procedures to estimate the original offspring distribution in the general framework provided by the CBPs. To this end, we make use of the minimum disparity methodology. This methodology has arisen as one that attains robustness properties without loss of efficiency. It was introduced in [30] for discrete models and since then, the literature on it has experimented a large growth (see [2, 32] for further information). In our context, assuming that the offspring distribution belongs to a very general parametric family, we determine minimum disparity estimators (MDEs) of the underlying parameter and study their asymptotic and robustness properties. The method consists in minimizing the discrepancy between a nonparametric estimator of the offspring distribution and the considered parametric family. The discrepancy is measured by a function called disparity measure. Thus, one can obtain different MDEs depending on the nonparametric estimator and the disparity measure considered. Special interest is highlighted in this paper for the negative exponential disparity and the Hellinger distance. The maximum likelihood estimator based on the observation of the whole family tree until a certain generation is considered as the nonparametric estimator. This paper presents for the first time the application of the technique of minimum disparity for the general class of branching structure given by CBPs, hence extending the results in [42] in a double sense: model and measure, and moreover extending the results in [1, 33] from an independent and identically distributed (i.i.d.) and continuous context to a dependent and discrete setup. It is worthwhile to point out the fundamental roles played by the nonparametric estimator and the dependence structure of the CBP to obtain the asymptotic properties of the MDEs proposed, that require a different approach from those already established in the i.i.d. setting.

Besides the introduction, this paper is organized into 6 sections and an appendix. In Section 2, we present the formal model and establish some hypotheses that we assume throughout the paper. Section 3 is devoted to defining and describing minimum disparity estimation. The asymptotic properties of MDEs are also studied; to this end, we introduce the disparity functional associated to a disparity measure and research its properties. The robustness of MDEs is studied in Section 5. To illustrate this methodology, we present a simulated example in Section 6. Concluding remarks about the contributions of the paper are presented in Section 7. Finally, we dedicate an appendix to the proofs of the theorems, in order to facilitate the reading of the paper.

2. The probability model

We consider a *controlled branching process with random control function* (CBP). Mathematically, this process is a discrete–time stochastic model $\{Z_n\}_{n\in\mathbb{N}}$ defined recursively as:

$$Z_0 = N, \quad Z_{n+1} = \sum_{j=1}^{\phi_n(Z_n)} X_{nj}, \quad n \in \mathbb{N}_0, \tag{2.1}$$

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $N \in \mathbb{N}_0$, $\{X_{nj} : n = 0, 1, \ldots; j = 1, 2, \ldots\}$ and $\{\phi_n(k) : n, k = 0, 1, \ldots\}$ are two independent families of nonnegative integer valued random variables defined on the same probability space $(\Omega, \mathcal{A}, P)$. Moreover, X_{nj}, $n = 0, 1, \ldots$, $j = 1, 2, \ldots$, are i.i.d. random variables and for each $n = 0, 1, \ldots$, $\{\phi_n(k)\}_{k \geq 0}$, are independent stochastic processes with equal one–dimensional probability distributions. The empty sum in (2.1) is considered to be 0. We denote by $p = \{p_k\}_{k \geq 0}$ the common probability distribution of the random variables X_{nj}, i.e., $p_k = P[X_{nj} = k]$, $k \geq 0$, which is known as offspring distribution or reproduction law, and by m and σ^2 its mean and variance (assumed finite), and we referred to them as offspring mean and variance, respectively. We also denote the mean and the variance of the control variables by $\varepsilon(k) = E[\phi_0(k)]$ and $\sigma^2(k) = Var[\phi_0(k)]$, $k \in \mathbb{N}_0$ (assumed finite too).

Intuitively, Z_n denotes the number of individuals in generation n, and X_{ni} the number of offspring of the ith individual in generation n. With the control mechanism $\phi_n(\cdot)$, if $\phi_n(Z_n) < Z_n$ then $Z_n - \phi_n(Z_n)$ individuals are removed from the population (this can model an emigration process, the culling process in the ecological example above or the presence of predators), and therefore do not participate in the future evolution of the process. If $\phi_n(Z_n) > Z_n$ then $\phi_n(Z_n) - Z_n$ new individuals of the same type are added to the population participating under the same conditions as the others (this can model an immigration process or a re-population). No control is applied to the population when $\phi_n(Z_n) = Z_n$.

Particular cases of CBPs commented in the introduction can be obtained by considering the following specific control variables:

(a) The evolution of invasive animal species can be described by a CBP by considering $\phi_n(k) = \ell_{\text{inf}} I_{(0,\ell_{\text{inf}})}(k) + k I_{[\ell_{\text{inf}},\ell_{\text{sup}}]}(k) + \ell_{\text{sup}} I_{(\ell_{\text{sup}},\infty)}(k)$, with I_B standing for the indicator function of the set B, and ℓ_{inf} and ℓ_{sup} nonnegative numbers, $\ell_{\text{inf}} < \ell_{\text{sup}}$.

(b) The evolution of a population threatened by the presence of predators can be modeled by setting that $\phi_n(k)$ follows a binomial distribution with parameters k and γ, for each $n, k \in \mathbb{N}_0$, where γ represents the survival probability of an individual or by a Poisson distribution of parameter γk as γ is low and the offspring mean is high.

(c) The BGW process is obtained by considering $\phi_n(k) = k$, for each $n, k \in \mathbb{N}_0$.

(d) The branching process with immigration can be seen as a particular case of the CBP. One only needs to set $\phi_n(k) = k + Y_n$, for each $n, k \in \mathbb{N}_0$, where Y_n are $\mathbb{N}_0$-valued i.i.d. random variables which are also independent of X_{nj}, $n = 0, 1, \ldots$, $j = 1, 2, \ldots$.

(e) A branching process with random migration can be described as a CBP by considering $\phi_n(k) = \max\{0, k + M_n\}$ with $\{M_n\}_{n \geq 0}$ a sequence of i.i.d. random variables such that

$$P[M_n = 0] = p, \quad P[M_n = -1] = q, \quad P[M_n = 1] = r,$$

with $p + q + r = 1$, $p, q, r \in (0, 1)$.

(f) The branching process with immigration at the state 0 (see [7]) can be obtained by using the control function $\phi_n(k) = \max(1, k)$, for each $n, k \in \mathbb{N}_0$.

It is easy to verify that $\{Z_n\}_{n\geq 0}$ is a Markov chain with stationary transition probabilities. Moreover, from now on we assume

(a) $p_0 > 0$ or $P[\phi_n(k) = 0] > 0$, $k > 0$,
(b) $\phi_n(0) = 0$ almost surely $(a.s.)$.

Such conditions guarantee that 0 is an absorbing state and the states $k = 1, 2, \dots$ are transient. Whence it is verified that $P[Z_n \to 0] + P[Z_n \to \infty] = 1$.

In addition, for our purpose, we suppose that the offspring distribution belongs to a general parametric family:

$$\mathcal{F}_\Theta = \{p(\theta) : \theta \in \Theta\}, \tag{2.2}$$

where $p(\theta) = \{p_k(\theta)\}_{k\geq 0}$ and Θ is a subset of $\mathbb{R}$, that is, $p = p(\theta_0)$ for some $\theta_0 \in \Theta$, referred as to the offspring parameter. For ease of presentation, we establish the results for a scalar parameter θ, although these can be generalized for a vector value parameter. The aim of this paper is to estimate θ_0 efficiently and robustly by choosing $\theta \in \Theta$ which provides the best adjustment to the observed sample in terms of the disparity measures.

To develop this methodology we need to consider nonparametric estimators of the offspring distribution. In this sense, in [18], nonparametric maximum likelihood estimators (MLEs) based on different samples are provided. Let denote a generic nonparametric estimator of p based on a sample, say $\mathcal{X}_n$, by $\tilde{p}_n = \{\tilde{p}_{n,k}\}_{k\geq 0}$, satisfying $\tilde{p}_{n,k} \geq 0$, for each $k \geq 0$, and $\sum_{k=0}^{\infty} \tilde{p}_{n,k} = 1$ (where n indicates that we observe the data up to the generation n).

3. Minimum disparity estimation

In this section, we introduce the notions of disparity measure and minimum disparity estimator, and present several interesting examples of them. Although we focus our attention on probability distributions defined on the nonnegative integers, that is, those which can be offspring distributions, the definitions and results given in this section keep valid for whatever discrete model. Let Γ be the set of all probability distributions defined on the nonnegative integers, $\mathcal{F}_\Theta$ the parametric family introduced in (2.2), and $G(\cdot)$ a three times differentiable and strictly convex function on $[-1, \infty)$ with $G(0) = 0$. The *disparity measure* ρ_G corresponding to $G(\cdot)$ is defined for any $q \in \Gamma$ and $\theta \in \Theta$, as

$$\begin{aligned}
\rho_G : \Gamma \times \Theta &\to [0, \infty] \\
(q, \theta) &\mapsto \rho_G(q, \theta) = \sum_{k=0}^{\infty} G(\delta(q, \theta, k)) p_k(\theta),
\end{aligned}$$

where $\delta(q,\theta,k)$ denotes the *"Pearson residual at k"*, that is,

$$\delta(q,\theta,k) = \begin{cases} \frac{q_k}{p_k(\theta)} - 1, & \text{if } p_k(\theta) > 0; \\ 0, & \text{otherwise.} \end{cases}$$

Note that one has different disparity measures ρ_G by considering different functions $G(\cdot)$; however, we drop G and write simply ρ in order to ease the notation. Moreover, notice that the Pearson residual at k depends on the probability distribution q and on the parameter θ, and that $\delta(q,\theta,k) \in [-1,\infty)$, for each $q \in \Gamma$, $\theta \in \Theta$, and $k \geq 0$.

Due to the fact that $G(\cdot)$ is strictly convex, one has that ρ is nonnegative. Moreover, when $G(\cdot)$ is also nonnegative and has a unique zero at 0 it is verified that $\rho(q,\theta) = 0$ if and only if $q = p(\theta)$. Given a sample $\mathcal{X}_n$ and a nonparametric estimator of p, $\tilde{p}_n$, based on it, we define the *minimum disparity estimator (MDE) of* θ_0 for the disparity measure ρ based on $\tilde{p}_n$ as

$$\tilde{\theta}_n^{\rho}(\tilde{p}_n) = \arg\min_{\theta\in\Theta} \rho(\tilde{p}_n,\theta). \tag{3.1}$$

It is important to mention that this estimator might not exist unless we assume some regularity conditions on the disparity measure ρ and on the parametric space Θ. This issue is addressed below.

Remark 1. *Some interesting cases of nonnegative disparity measures are the following:*

(a) The disparity obtained with the function $G(\delta) = (\delta+1)\log(\delta+1)$ is a kind of the Kullback–Leibler divergence. It is denoted by $LD(\tilde{p}_n,\theta)$, for each $\tilde{p}_n$, $n \in \mathbb{N}$, and $\theta \in \Theta$, and it is known as likelihood disparity*. Its minimizer, $\tilde{\theta}_n^{LD}(\tilde{p}_n)$, is known as the minimum likelihood disparity estimator (MLDE). In some cases, this estimator coincides with the MLE.*

(b) The disparity determined by the function $G(\delta) = [(\delta+1)^{1/2} - 1]^2$ is the squared Hellinger distance*, denoted by $HD(\tilde{p}_n,\theta)$, for each $\tilde{p}_n$, $n \in \mathbb{N}$, and $\theta \in \Theta$. Notice that the squared Hellinger distance between two probability distributions is the square of the l_2-distance between the square roots of the corresponding probability distributions, i.e., $HD(q,\theta) = \|q^{1/2} - p(\theta)^{1/2}\|_2^2$, where $\|\cdot\|_2$ denotes the l_2-norm defined on Γ, and for each $q \in \Gamma$, $q^{1/2} = \{q_k^{1/2}\}_{k\geq 0}$. In this case, the MDE is the minimum Hellinger distance estimator (MHDE), denoted by $\tilde{\theta}_n^{HD}(\tilde{p}_n)$.*

(c) The disparities defined by using either the function $G(\delta) = \exp(-\delta) - 1$ or $G(\delta) = \exp(-\delta) - 2$ (denoted by $D(\tilde{p}_n,\theta)$ and $D_M(\tilde{p}_n,\theta)$, respectively, for each $\tilde{p}_n$, $n \in \mathbb{N}$, and $\theta \in \Theta$) are both known as negative exponential disparity *(notice both disparities differ only in a constant). The MDE is denoted by $\tilde{\theta}_n^{NED}(\tilde{p}_n)$ and it is called the minimum negative exponential disparity estimator (MNEDE).*

Other examples are the family of power divergence measures *(see [8]), the* blended chi-squared measures*, which include* Pearson's chi–squared *and* Neyman's chi-square*, the* blended weight chi–squared measures *and the* blended weight Hellinger distance family *(see [30]).*

Under conditions of differentiability of the model, a useful way for determining a MDE is to take into account that it must satisfy $\dot{\rho}(\tilde{p}_n, \tilde{\theta}_n^\rho(\tilde{p}_n)) = 0$, where for each $q \in \Gamma$, $\dot{\rho}(q, \theta)$ denotes the first derivative of $\rho(q, \theta)$ with respect to θ. It is verified, for $q \in \Gamma$ and $\theta \in \Theta$,

$$-\dot{\rho}(q, \theta) = \sum_{k=0}^{\infty} p_k'(\theta) A(\delta(q, \theta, k)),$$

where $A(\delta) = (\delta+1)G'(\delta) - G(\delta)$, and $G'(\cdot)$ and $p_k'(\cdot)$ denote the first derivative of $G(\cdot)$ and $p_k(\cdot)$, respectively. The function $A(\cdot)$ is called the *residual adjustment function* (RAF) of the disparity. It is twice differentiable and an increasing function on $[-1, \infty)$ that can be redefined (standarized), without changing the estimating properties of the disparity, so that $A(0) = 0$, and $A'(0) = 1$.

Remark 2. *The function $G(\cdot)$ and the RAF of the disparity measures of Remark 1 after standardization (that is, for which $A(0) = 0$ and $A'(0) = 1$) are the following:*

- *(a) The RAF of the likelihood disparity is $A(\delta) = \delta$ and $G(\delta) = (\delta+1)\log(\delta+1) - \delta$.*
- *(b) The RAF of the squared Hellinger distance is $A(\delta) = 2[(\delta+1)^{1/2} - 1]$, corresponding to the function $G(\delta) = 2((\delta+1)^2 - 1)^2$.*
- *(c) The RAF of the negative exponential disparity is $A(\delta) = -(2+\delta)\exp(-\delta) + 2$, corresponding to $G(\delta) = \exp(-\delta) - 1 + \delta$.*

In [30], the RAFs of different disparity measures are compared (see Figures 4 and 5 in [30]). The RAF of a disparity measure is relevant in determining the efficiency and robustness properties of the corresponding MDE. Concretely, $A''(0)$ is demonstrated to play a key role: large negative values of $A''(0)$ correspond to robustness properties and zero value matches a second–order efficient estimator in the sense of [37].

Before focusing on these matters, first we establish the existence of the minimum in (3.1) and its uniqueness. To this end, the approach followed is to view $\tilde{\theta}_n^\rho(\tilde{p}_n)$ as the value of a functional T^ρ at $\tilde{p}_n$. We consider the *disparity functional* associated with a disparity measure ρ, defined as $T^\rho : \Gamma \to \Theta$, with $T^\rho(q) = \arg\min_{\theta\in\Theta} \rho(q, \theta)$, whenever the minimum exists. Notice that there might exist multiple values of the parameter θ which minimize the function $\rho(q, \cdot)$. As a consequence, $T^\rho(q)$ would denote any of these values. Clearly, $T^\rho(\tilde{p}_n) = \tilde{\theta}_n^\rho(\tilde{p}_n)$.

It is obvious that if the parameter space Θ is compact and the function $\rho(q, \cdot)$ is continuous in Θ, for each $q \in \Gamma$, then $T^\rho(q)$ exists. However, we will weaken the compactness of Θ in a similar way as was done in [39] and in [42]. Specifically, given a disparity ρ, we limit our study to the subclass of probability distributions $\tilde{\Gamma}_\rho \subseteq \Gamma$ which satisfies the following condition: there exists a compact set $C_\rho \subseteq \Theta$ such that for every $q \in \tilde{\Gamma}_\rho$,

$$\inf_{\theta\in\Theta\setminus C_\rho} \rho(q, \theta) > \rho(q, \theta^*), \tag{3.2}$$

for some $\theta^* \in C_\rho$.

Theorem 3.1 (Existence)**.** *It is satisfied:*

(i) *For each $q \in \tilde{\Gamma}_\rho$ satisfying that $\rho(q,\cdot)$ is continuous in C_ρ, there exists $T^\rho(q)$.*

(ii) *If ρ is a disparity measure, $\theta^* \in \Theta$ verifies $\inf_{\theta\in\Theta\setminus K}\rho(p(\theta^*),\theta) > 0$ for some compact set $K \subseteq \Theta$ and $\rho(p(\theta^*),\cdot)$ is continuous in K, then $T^\rho(p(\theta^*))$ exists. Moreover, if $\mathcal{F}_\Theta$ is identifiable, that is, $p(\cdot)$ is injective, and the disparity ρ can be redefined (without changing its minimizer) so that the related function $G(\cdot)$ is nonnegative and has a unique zero at 0, then $\theta^* = T^\rho(p(\theta^*))$.*

The proof is provided in Appendix.

Remark 3. *(a) Notice that the continuity of $p_k(\cdot)$ in an arbitrary set $B \subseteq \Theta$ for each $k \geq 0$ leads to the continuity in B of the function $\rho(q,\cdot)$ associated with any disparity measure ρ determined by a bounded function $G(\cdot)$, for each $q \in \Gamma$. This is deduced from a generalized dominated convergence theorem (see [38], p.92). The aforesaid condition is satisfied by the negative exponential disparity. Although the Hellinger distance is defined by a non bounded function $G(\cdot)$, in this case the condition of continuity of $p_k(\cdot)$ in B for each $k \geq 0$ is enough to obtain the continuity of $HD(q,\cdot)$ in B for each $q \in \Gamma$. This latter is followed by the Cauchy-Schwarz inequality and the Scheffé's theorem.*

(b) The redefinition of some disparities, without affecting their minimizer, so that the related functions $G(\cdot)$ will be nonnegative and have a unique zero at 0 is possible. For instance, for the negative exponential disparity we can consider the function $\bar{G}(\delta) = G(\delta) + \delta$ instead of $G(\delta)$, which verifies the previous properties and for each $q \in \Gamma$, $\sum_{k=0}^\infty \bar{G}(\delta(q,\theta,k))p_k(\theta) = \sum_{k=0}^\infty G(\delta(q,\theta,k))p_k(\theta)$.

In order to study the asymptotic properties of the MDEs, we have to assume several conditions. Let fix the next assumptions:

(A1) ρ is a disparity measure associated with a function $G(\cdot)$ which satisfies that $G(\cdot)$ and $G'(\cdot)$ are bounded in $[-1,\infty)$.

(A2) $p_k(\cdot)$ is continuous in C_ρ for each $k \geq 0$ (where C_ρ is introduced in (3.2)).

Notice that under (A2), by Theorem 3.1, $T^\rho(q)$ exists for every $q \in \tilde{\Gamma}_\rho$. Let $\hat{\Gamma}_\rho$ be the set of $q \in \tilde{\Gamma}_\rho$ such that $T^\rho(q)$ is unique. Now, in the following theorem the continuity of the disparity functional is established. Henceforth, all the limits are taken as $n \to \infty$.

Theorem 3.2 (Continuity)**.** *Let q and $\{q_n\}_{n\in\mathbb{N}}$ be in Γ such that $q_n \to q$ in l_1. Assuming (A1), (A2) and that $q \in \hat{\Gamma}_\rho$, then $T^\rho(q_n)$ eventually exists and the functional $T^\rho(\cdot)$ is continuous in q, that is, $T^\rho(q_n) \to T^\rho(q)$.*

Analogously, since the Hellinger distance does not satisfy condition (A1), under the following alternative hypotheses, the result is also established.

Theorem 3.3 (Continuity for the Hellinger distance)**.** *Let q and $\{q_n\}_{n\in\mathbb{N}}$ be in Γ satisfying $||q_n^{1/2} - q^{1/2}||_2 \to 0$. If (A2) holds and $q \in \hat{\Gamma}_{HD}$, then $T^{HD}(q_n)$ eventually exists and the functional $T^{HD}(\cdot)$ is continuous in q, that is, $T^{HD}(q_n) \to T^{HD}(q)$.*

The proofs of Theorems 3.2 and 3.3 are given in Appendix.

Recall that $p = p(\theta_0)$ is the true reproduction law. Observe that under (A1), if (A2) is verified and $p \in \hat{\Gamma}_\rho$, one obtains $T^\rho(p) = \theta_0$ and for the case of Hellinger distance, dropping (A1), one also has $T^{HD}(p) = \theta_0$. Next theorem establishes the strong consistency of the MDEs.

Theorem 3.4 (Consistency)**.** *Assume (A2) and $p \in \hat{\Gamma}_\rho$, for the corresponding disparity ρ. Then if $\tilde{p}_{n,k}$ converges to $p_k(\theta_0)$ a.s., for each $k \geq 0$, one has that:*

(i) $\tilde{\theta}_n^\rho(\tilde{p}_n)$ eventually exists, is a random variable, and $\tilde{\theta}_n^\rho(\tilde{p}_n) \to \theta_0$ a.s. if (A1) holds.
(ii) $\tilde{\theta}_n^{HD}(\tilde{p}_n)$ eventually exists, is a random variable, and $\tilde{\theta}_n^{HD}(\tilde{p}_n) \to \theta_0$ a.s.

The proof can be consulted in Appendix.

4. Asymptotic normality

The results given in the previous section are general in the sense that the explicit expression of the nonparametric estimator is not required, and one only needs to know its properties, as for example, its consistency. However, to establish the asymptotic normality of the MDEs, explicit formulas of the nonparametric estimators are needed. For this reason, to develop this section we come back to the CBP context.

In [18], we provide nonparametric estimators of the offspring distribution under several sampling schemes. In particular, in a complete data context, we consider that the entire family tree up to generation n can be observed, that is, the sample $\mathcal{Z}_n^* = \{Z_l(k) : 0 \leq l \leq n-1; k \geq 0\}$, where $Z_l(k) = \sum_{i=1}^{\phi_l(Z_l)} I_{\{X_{li}=k\}}$, $0 \leq l \leq n-1$, $k \geq 0$, with recall I_B standing for the indicator function of the set B. Intuitively, $Z_l(k)$ represents the number of parents in generation l who have exactly k offspring. Thus, it is easy to check that $\phi_l(Z_l) = \sum_{k=0}^\infty Z_l(k)$, and $Z_{l+1} = \sum_{k=0}^\infty kZ_l(k)$, for $l \in \mathbb{N}_0$. Recall that in a general setting $p = \{p_k\}_{k\geq 0}$ is the offspring distribution. The MLE of p_k, for each $k \geq 0$, (see [18]), is given by $\widehat{p}_n = \{\widehat{p}_{n,k}\}_{k\geq 0}$:

$$\widehat{p}_{n,k} = \frac{Y_{n-1}(k)}{\Delta_{n-1}}, \qquad k \geq 0, \tag{4.1}$$

where $\Delta_l = \sum_{j=0}^l \phi_j(Z_j)$, and $Y_l(k) = \sum_{j=0}^l Z_j(k)$, $k \geq 0$, $0 \leq l \leq n-1$. Intuitively, Δ_l is the total number of parents until generation l, and $Y_l(k)$ is the total number of progenitors with exactly k offspring until generation l. Consequently, one estimates the probability that an individual has k offspring as the relative proportion of parents with k offspring. For the BGW process (recall $\phi_n(k) = k$ for each k and n), $\widehat{p}_n$ corresponds to the estimator given in [24], p. 42.

It is proved that $\widehat{p}_{n,k}$ is strongly consistent for p_k on $\{Z_n \to \infty\}$, for each $k \geq 0$, (see Theorem 3.6 in [18]), under the following assumption:

(A3) The CBP satisfies that:

(a) There exists $\tau = \lim_{k\to\infty} \varepsilon(k)k^{-1} < \infty$, and the sequence $\{\sigma^2(k)k^{-1}\}_{k\geq 1}$ is bounded.

(b) $\tau_m = \tau m > 1$ and Z_0 is large enough such that $P[Z_n \to \infty] > 0$.

(c) $\{Z_n \tau_m^{-n}\}_{n\geq 0}$ converges a.s. to a finite random variable W such that $P[W > 0] > 0$.

(d) $\{W > 0\} = \{Z_n \to \infty\}$ a.s.

In order to establish the asymptotic normality of the MDEs, we add to assumption (A3)(c) that $\{Z_n \tau_m^{-n}\}_{n\geq 0}$ converges in L^1 to a finite random variable W, with $0 < E[W] < \infty$.

Remark 4. *(a) For CBPs verifying (A3)(a), sufficient conditions for (A3)(b)-(d) are discussed in [19]–[22].*

(b) It is easy to check that $\tau_m = \lim_{k\to\infty} E[Z_{n+1}|Z_n = k]k^{-1}$. For each $k \in \mathbb{N}$, $E[Z_{n+1}|Z_n = k]k^{-1}$ can be interpreted as a mean growth rate, thus, τ_m is referred to as the asymptotic mean growth rate. This is the threshold parameter of CBPs (see [21]).

(c) It can be proved that, under (A3), $\phi_n(Z_n)Z_n^{-1} \to \tau$ a.s., and $\varepsilon(Z_n)/\phi_n(Z_n) \to 1$ a.s. on $\{Z_n \to \infty\}$. As a result, $\phi_n(Z_n)\tau_m^{-n} \to \tau W$ a.s. on $\{Z_n \to \infty\}$ (see [18], Proposition 3.5).

(d) Since the BGW process can be seen as a CBP with $\phi_n(k) = k$, for each $n, k \in \mathbb{N}_0$, one has that $\varepsilon(k) = k$ and $\sigma^2(k) = 0$, for each $k \in \mathbb{N}_0$. Consequently, (A3) is fulfilled provided that $m > 1$ (and taking into account that $\sigma^2 < \infty$).

As a consequence, under (A3), from Theorem 3.4 *(i)* and *(ii)*, one obtains, respectively, that the estimators $\tilde{\theta}_n^{\rho}(\hat{p}_n)$ and $\tilde{\theta}_n^{HD}(\hat{p}_n)$ are strongly consistent on $\{Z_n \to \infty\}$.

Now, we focus our attention on the asymptotic normality. To this end, we must consider additional conditions on the functions $p(\cdot)$. From now on, we assume that for each $k \geq 0$, $p_k(\theta)$ is twice continuously differentiable with respect to θ and:

(A4) For $\theta \in \Theta$, $\epsilon > 0$, for each $\theta^* \in (\theta - \epsilon, \theta + \epsilon)$, and $k \geq 0$,

(a) there exists $J_k(\theta)$ such that $|p_k'(\theta^*)| < J_k(\theta)$, and $\sum_{k=0}^{\infty} J_k(\theta) < \infty$,

(b) there exists $L_k(\theta)$ such that $|p_k''(\theta^*)| < L_k(\theta)$, and $\sum_{k=0}^{\infty} L_k(\theta) < \infty$,

(c) there exists $M_k(\theta)$ such that $|u(\theta^*, k)^2 p_k(\theta^*)| < M_k(\theta)$, and $\sum_{k=0}^{\infty} M_k(\theta) < \infty$, where $u(\theta, k) = (\log(p_k(\theta)))' = p_k'(\theta)/p_k(\theta)$.

(A5) ρ is a disparity measure whose RAF $A(\cdot)$ satisfies that $A(0) = 0$, $A'(0) = 1$, and $A(\delta)$, $A'(\delta)$, $A'(\delta)(1 + \delta)$ and $A''(\delta)(1 + \delta)$ are bounded functions on $\delta \in [-1, \infty)$.

Remark 5. *Notice that for a disparity measure ρ satisfying (A5), (A4) is a sufficient condition to guarantee that $\rho(q,\theta)$ can be twice differentiable with respect to θ.*

It is easy to check that the negative exponential disparity satisfies (A5) but the Hellinger distance does not. In the latter case, to establish the efficiency of MHDE, instead of the previous hypotheses, we assume the following condition on $s(\theta)=\{s_k(\theta)\}_{k\geq 0}$, with $s_k(\theta)=p_k(\theta)^{1/2}$, in a similar way to that in [3]:

(A6) For $\theta\in int(\Theta)$ (that is, θ in the interior of Θ), $s(\theta)$ is twice differentiable in l_2; that is, there exist $s'(\theta)=\{s'_k(\theta)\}_{k\geq 0}\in l_2$ and $s''(\theta)=\{s''_k(\theta)\}_{k\geq 0}\in l_2$ and for every β in a neighbourhood of zero

$$s_k(\theta+\beta)=s_k(\theta)+\beta s'_k(\theta)+\beta v_k(\beta),$$

$$s'_k(\theta+\beta)=s'_k(\theta)+\beta s''_k(\theta)+\beta w_k(\beta),$$

where $\sum_{k=0}^{\infty}v_k(\beta)^2\to 0$, and $\sum_{k=0}^{\infty}w_k(\beta)^2\to 0$, as $\beta\to 0$.

Let denote $I(\theta_0)=\sum_{k=0}^{\infty}u(\theta_0,k)^2p_k(\theta_0)$, the Fisher information for θ contained in the random variable X_{01}. Since $I(\theta_0)=4||s'(\theta_0)||_2^2$, either from (A4)(c) or from $s'(\theta_0)\in l_2$, $I(\theta_0)<\infty$ is obtained. In addition, observe that although conditions (A1) and (A5) seem to be quite restrictive, they are satisfied by a wide set of disparities (see [33]).

Theorem 4.1 (Asymptotic normality)**.** *Let be a CBP satisfying (A3), with $p=p(\theta_0)$ its offspring distribution. Moreover, assume (A2) and $p\in\hat{\Gamma}_\rho$ (recall that in this case $T^\rho(p)=\theta_0$).*

(i) If (A1), (A4), and (A5) hold, $s'(\theta_0)\in l_1$, and supposing that any sequence of estimators $\{\varphi_n\}_{n\in\mathbb{N}}$ converging to θ_0 in probability satisfies

$$\sum_{k=0}^{\infty}|p''_k(\varphi_n)-p''_k(\theta_0)| \quad \xrightarrow{P} \quad 0, \tag{4.2}$$

$$\sum_{k=0}^{\infty}|u(\varphi_n,k)^2p_k(\varphi_n)-u(\theta_0,k)^2p_k(\theta_0)| \quad \xrightarrow{P} \quad 0, \tag{4.3}$$

then, it is verified:

$$\Delta_{n-1}^{1/2}(\tilde{\theta}_n^\rho(\hat{p}_n)-\theta_0)\xrightarrow{d} N\left(0,I(\theta_0)^{-1}\right), \tag{4.4}$$

where $\xrightarrow{P}$ denotes the convergence in probability and $\xrightarrow{d}$ represents the convergence in distribution with respect to the probability $P[\cdot|Z_n\to\infty]$.

(ii) For the Hellinger distance, (4.4) also holds under the assumptions (A6), $\theta_0\in int(\Theta)$, and $\sum_{k=0}^{\infty}s''_k(\theta_0)p_k^{1/2}<0$.

The proof of the previous theorem is given in Appendix.

Remark 6. *(a) Notice that to establish the asymptotic normality of the MDEs, the assumptions are imposed on the offspring distribution and on the disparity measure, so that taking into account Remark 4, (c), for a BGW process one has that under hypotheses in Theorem 4.1,*

$$\left(\sum_{j=0}^{n-1} Z_j\right)^{1/2} (\tilde{\theta}_n^\rho(\hat{p}_n) - \theta_0) \xrightarrow{d} N\left(0, I(\theta_0)^{-1}\right).$$

(b) Assuming that we observe $\mathcal{Z}_n^*$, one knows $\phi_l(Z_l) = \sum_{k=0}^\infty Z_l(k)$ and $Z_{l+1} = \sum_{k=0}^\infty kZ_l(k)$, $l = 0, \ldots, n-1$. Consequently, using Theorem 3.4, 4.4, the continuity of $I(\theta)$, and Slutsky's theorem,

$$\tilde{\theta}_n^\rho(\hat{p}_n) \pm z_\gamma(\Delta_{n-1} I(\tilde{\theta}_n^\rho(\hat{p}_n))^{-1/2},$$

with z_γ the quantile of order $1-\gamma/2$ of a standard Normal distribution, provides an asymptotic confidence interval at $(1-\gamma)\%$ level.

(c) Besides the MDEs based on the whole family tree, one can determine the ones based on other samples. In [18], we also study the maximum likelihood estimation of the offspring distribution under incomplete sampling schemes, considering the random samples given by the number of individuals and progenitors in each generation, that is, $\overline{\mathcal{Z}}_n = \{Z_0, Z_{l+1}, \phi_l(Z_l) : l = 0, \ldots, n-1\}$, and only by the generation sizes, that is, $\mathcal{Z}_n = \{Z_0, \ldots, Z_n\}$. The proposed estimators for the offspring distribution, based on $\overline{\mathcal{Z}}_n$ and $\mathcal{Z}_n$, are obtained by the Expectation–Maximization algorithm (EM algorithm). Making use of these estimators, one can obtain MDEs of θ_0 based on $\overline{\mathcal{Z}}_n$ and $\mathcal{Z}_n$, respectively.

5. Robustness

In this section, we address the issue of the robustness of the MDEs of θ. For this purpose, we will study the behaviour of the corresponding disparity functional under contamination by considering the following model:

$$p(\alpha, \theta, L) = (1-\alpha)p(\theta) + \alpha\eta_L, \tag{5.1}$$

where $\alpha \in (0,1)$, $\theta \in \Theta$, $L \in \mathbb{N}_0$ and η_L is a point mass distribution at a nonnegative integer L. This model is called mixture model for gross errors at L and it represents the simplest context of contamination. This approach was introduced in [45] and consists in assuming the *contaminated model* instead of the model distribution in order to explain or incorporate outliers.

In the analysis of robustness of an estimator, an essential tool is the *influence curve*, which for each disparity ρ is a function of $L \in \mathbb{N}_0$ defined as

$$\lim_{\alpha \to 0} \alpha^{-1}(T^\rho(p(\alpha, \theta, L)) - T^\rho(p(\theta))).$$

Although the unboundedness of this function is an indicator of the misbehaviour of the MDEs of θ in presence of outliers, the influence curve can be very a deceptive measure of robustness (see [30]). For this reason, we will also examine the *α-influence curves of* $T^{\rho}(\cdot)$, which are functions of $L \in \mathbb{N}_0$ defined as $\alpha^{-1}(T^{\rho}(p(\alpha, \theta, L)) - T^{\rho}(p(\theta)))$, for each $\alpha \in (0, 1)$.

Next theorem, whose proof can be read in Appendix, provides an expression for the influence curves and establishes conditions under which the disparity functional is robust at $p(\theta)$ against $100\alpha\%$ contamination by gross error at an arbitrary integer.

Theorem 5.1 (Robustness)**.** *Suppose the parameter space Θ is compact, (A2) holds (with $C_\rho = \Theta$), and $\mathcal{F}_\Theta$ is identifiable. For every $\alpha \in (0, 1)$ and every $\theta \in \Theta$:*

(i) Let ρ be a disparity measure which can be redefined (without changing its minimizer) so that the related function $G(\cdot)$ is nonnegative, has a unique zero at 0 and satisfies (A1) and (A5). For each $q \in \Gamma$ and $t \in \Theta$, define $\rho^*(\alpha, q, t) = \sum_{k=0}^{\infty} G^*(\delta(q, t, k))p_k(t)$, with $G^*(\delta) = G((1-\alpha)\delta)$. If (A4) holds, $T^{\rho}(p(\theta, \alpha, L))$ is unique for all $L \in \mathbb{N}_0$, and there exists a strictly increasing function f such that $f(\rho^*(\alpha, p(\theta), t)) = \rho((1-\alpha)p(\theta), t)$, for $t \in \Theta$; then

(a) $\lim_{L\to\infty} T^{\rho}(p(\theta, \alpha, L)) = \theta$.

(b) $T^{\rho}(p(\theta, \alpha, L))$ *is a bounded and continuous function of* L.

(c) $\lim_{\alpha\to 0} \alpha^{-1}(T^{\rho}(p(\theta, \alpha, L)) - \theta) = (I(\theta)p_L(\theta))^{-1}p'_L(\theta)$.

(ii) For the Hellinger distance, if $T^{HD}(p(\theta)) \in int(\Theta)$, $\sum_{k=0}^{\infty} s''_k(\theta)p_k^{1/2} < 0$, (A6) holds, and $T^{HD}(p(\theta, \alpha, L))$ is unique for all $L \in \mathbb{N}_0$; then (i-a), (i-b) and (i-c) are also satisfied.

Observe that $p'_L(\theta)(I(\theta)p_L(\theta))^{-1}$ can be an unbounded function of L. Nevertheless, from Theorem 5.1 *(a)* and *(b)*, we have that for every $\alpha \in (0, 1)$, the α–influence curves are bounded continuous functions of L satisfying $\lim_{L\to\infty} \alpha^{-1} \cdot (T^{\rho}(p(\theta, \alpha, L)) - \theta) = 0$, and $\lim_{L\to\infty} \alpha^{-1}(T^{HD}(p(\theta, \alpha, L)) - \theta) = 0$, respectively. Consequently, the associated disparity functionals are robust at $p(\theta)$ against $100\alpha\%$ contamination by gross error at an arbitrary integer L.

Another important concept in the study of the robustness is the *asymptotic breakdown point*. Intuitively, the asymptotic breakdown point represents the smallest amount of contamination that can cause the estimator to take arbitrarily large values. Formally, the asymptotic breakdown point of a disparity functional $T^{\rho}(\cdot)$ at $q \in \Gamma$ is given by:

$$\alpha^*(T^{\rho}, q) = \inf \{\alpha \in (0, 1) : b(\alpha; T^{\rho}, q) = \infty\},$$

with $b(\alpha; T^{\rho}, q) = \sup \{|T^{\rho}((1-\alpha)q + \alpha\overline{q}) - T^{\rho}(q)| : \overline{q} \in \Gamma\}$. Note that $b(\alpha; T^{\rho}, q) = \infty$ is equivalent to the existence of a sequence of probability distributions $\{q_n\}_{n\in\mathbb{N}}$ satisfying $|T^{\rho}((1-\alpha)q + \alpha q_n) - T^{\rho}(q)| \to \infty$ and in that case, we say there is *breakdown* in $T^{\rho}(\cdot)$ for a level of contamination equals α. The sequence $\{q_n\}_{n\in\mathbb{N}}$ is called sequence of contaminating probability distributions. This fact is useful for the establishment of a lower bound for the asymptotic breakdown point of $T^\rho(\cdot)$ at some probability distribution in the following theorem, and in particular, for determining the asymptotic breakdown point of the MDEs of θ_0 based on nonparametric estimators of the offspring distribution.

Theorem 5.2 (Asymptotic breakdown point)**.**

(i) *Assume that the contaminant sequence $\{q_n\}_{n\in\mathbb{N}}$, the distributions of the family $\mathcal{F}_\Theta$, and $\theta^* \in \Theta$ satisfy:*

(a) $\sum_{k=0}^{\infty} \min\{p_k(\theta^*), q_{n,k}\} \to 0$.

(b) $\sum_{k=0}^{\infty} \min\{p_k(\theta), q_{n,k}\} \to 0$, *uniformly for $\theta \in \Theta$ such that $|\theta| \leq c$, for any fixed $c \in \mathbb{R}$.*

(c) $\sum_{k=0}^{\infty} \min\{p_k(\theta^*), p_k(\theta_n)\} \to 0$, *if $|\theta_n| \to \infty$.*

(d) $G(-1)$ *and* $\lim_{t\to\infty} G(t)/t$ *are finite.*

Then, the asymptotic breakdown point of the MDE of θ^* is at least $1/2$.

(ii) *Assume (A2), $q \in \hat{\Gamma}_{HD}$, and $T^{HD}(q) \in int(\Theta)$. Let $\varrho(q, p(\theta)) = \sum_{k=0}^{\infty}(q_k p_k(\theta))^{1/2}$, $\hat{\varrho} = \max_{\theta\in\Theta} \varrho(q, p(\theta))$, $\varrho^* = \lim_{M\to\infty} \sup_{|\theta|>M} \varrho(q, p(\theta))$ and $h_n = (1-\alpha)q + \alpha q_n$, $0 < \alpha < 1$, with $q_n \in \Gamma$, for every n. Assume that for each $n \geq 1$, $T^{HD}(h_n)$ exists and is unique. It holds that if $\alpha < (\hat{\varrho} - \varrho^*)^2/[1 + (\hat{\varrho} - \varrho^*)^2]$, then there is no sequence $\{h_n\}_{n\in\mathbb{N}}$ of the form defined above for which $\lim_{n\to\infty} |T^{HD}(h_n) - T^{HD}(q)| = \infty$.*

The proof of *(i)* is analogous to that given in Theorem 4.1 in [33] replacing integrals with sums. Intuitively, the assumptions *(a)*-*(c)* indicate the worst possible selection of the contamination and represent the asymptotic singularity between the probability distributions considered (see [33] for a further description). The proof of *(ii)* is exactly the same as Theorem 3 in [39] and it is omitted. In particular, when q is the offspring distribution $p = p(\theta_0)$, then $\hat{\varrho} = 1$ and as a consequence, the asymptotic breakdown point for HD is at least $1/2$ when $\varrho^* = 0$, which usually holds.

6. Simulated example

Through a simulated example, we compare the behaviour of the MHDEs, MNEDEs and MLDEs based on the whole family tree under an uncontaminated model and under mixture models for gross errors. To this end, we have considered as initial model a CBP starting with one individual and Poisson distributions as offspring and control distributions. In particular, the offspring distribution is a Poisson distribution with the parameter $\theta_0 = 7$ and the variable $\phi_n(k)$ follows Poisson distribution with parameter λk, with $\lambda = 0.3$, for each $k \geq 0$, $n \geq 0$. Therefore, the offspring mean and variance are $m = \sigma^2 = 7$, and $\tau_m = \theta_0\lambda = 2.1$ (see (A3) for definition). In practice, control functions $\phi_n(k)$ that follow Poisson distributions with parameters λk are appropriate to describe an environment with expecting immigration or emigration according to the value of the parameter λ: the former corresponds to $\lambda > 1$ and the latter to $\lambda < 1$. Recall as was pointed out in the introduction, this control distribution can be used to model populations threatened by a predator as an approximation to the binomial control process. In our example, we consider a model with expected emigration although supercritical ($\tau_m > 1$).

First, we show that in a contamination-free context, MHDEs and MNEDEs are as efficient as MLDEs. To this end, we have simulated 10 generations of $N = 100$ CBPs following the previous model, and we have estimated the relative efficiency of $\tilde{\theta}_n^{NED}(\hat{p}_n)$ to $\tilde{\theta}_n^{HD}(\hat{p}_n)$, of $\tilde{\theta}_n^{HD}(\hat{p}_n)$ to $\tilde{\theta}_n^{LD}(\hat{p}_n)$ and of $\tilde{\theta}_n^{NED}(\hat{p}_n)$ to $\tilde{\theta}_n^{LD}(\hat{p}_n)$ in each generation by the ratios of these mean squared errors:

$$\frac{\mathrm{MSE}(HD)}{\mathrm{MSE}(NED)}, \ \frac{\mathrm{MSE}(LD)}{\mathrm{MSE}(HD)}, \ \frac{\mathrm{MSE}(LD)}{\mathrm{MSE}(NED)},$$

where MSE(ρ)= $N^{-1}\sum_{i=1}^{N}(\tilde{\theta}_i^{\rho}(\hat{p}_n) - \theta_0)^2$, $\rho \in \{LD,\ NED,\ HD\}$, where n indicates the generation, for $n = 1, \ldots, 10$, and i indicates the simulated process, for $i = 1, \ldots, N$. The evolution of these estimates is shown in Figure 1 (first row -left), where one observes that as generations go up, MNED and MHD procedures are shown as efficient as the MLD one.

In a contaminated context, to illustrate and compare the accuracy of the estimates obtained by MHD and MNED methods, we have considered several different contaminated models for the offspring distribution in the aforementioned CBP. Specifically, we have contaminated the reproduction law according to the mixture model for gross errors, for $\alpha = 0.05$, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, and $L = 0, 1, \ldots, 25$, obtaining 260 different contaminated CBPs.

For a generic CBP, the information given by a sample observed until a fixed generation n depends on its asymptotic mean growth rate, τ_m, and it is poorer when $\tau_m \approx 1$ than when $\tau_m > 1$. This implies that to compare the behaviour of the different estimators for each one of the contaminated models (which have asymptotic mean growth rates, called $\tau_m(\theta_0, \alpha, L)$, of different magnitudes) one needs to observe different numbers of generations depending upon the value of $\tau_m(\theta_0, \alpha, L)$. In our example, these values go from $n = 8$ for $\tau_m(\theta_0, \alpha, L) = 4.8$ to $n = 65$ for $\tau_m(\theta_0, \alpha, L) = 1.05$.

For each simulated process, we have determined the MHDEs, MNEDEs and MLDEs of θ_0 in its last generation. In Figure 1, we show the mean (over the 100 simulations) of the MHDEs (first row -right) and of the MNEDEs (second row -left) of θ_0, for each one of the 260 contaminated models. Moreover, the respective MSEs for both methods are represented in Figure 1 (second row -right and third row -left).

In addition, Figure 1 (third row -right) shows the contour plot of the asymptotic mean growth rate of the contaminated models, $\tau_m(\theta_0, \alpha, L)$, and the underlying points represent the minimum disparity method which provides the smallest MSE for each contaminated model. There are two remarkable facts that can be deduced from this plot. The first one is that the MNEDE supplies more accurate estimates in most of the contaminated models (166 models, that is 63.85% of the models), but the best method when the contaminated state is between 3 and 11 is usually the MHD (85 models, that is, 32.69% of models). Moreover, the MLDE only behaves properly in 9 models (3.46% of

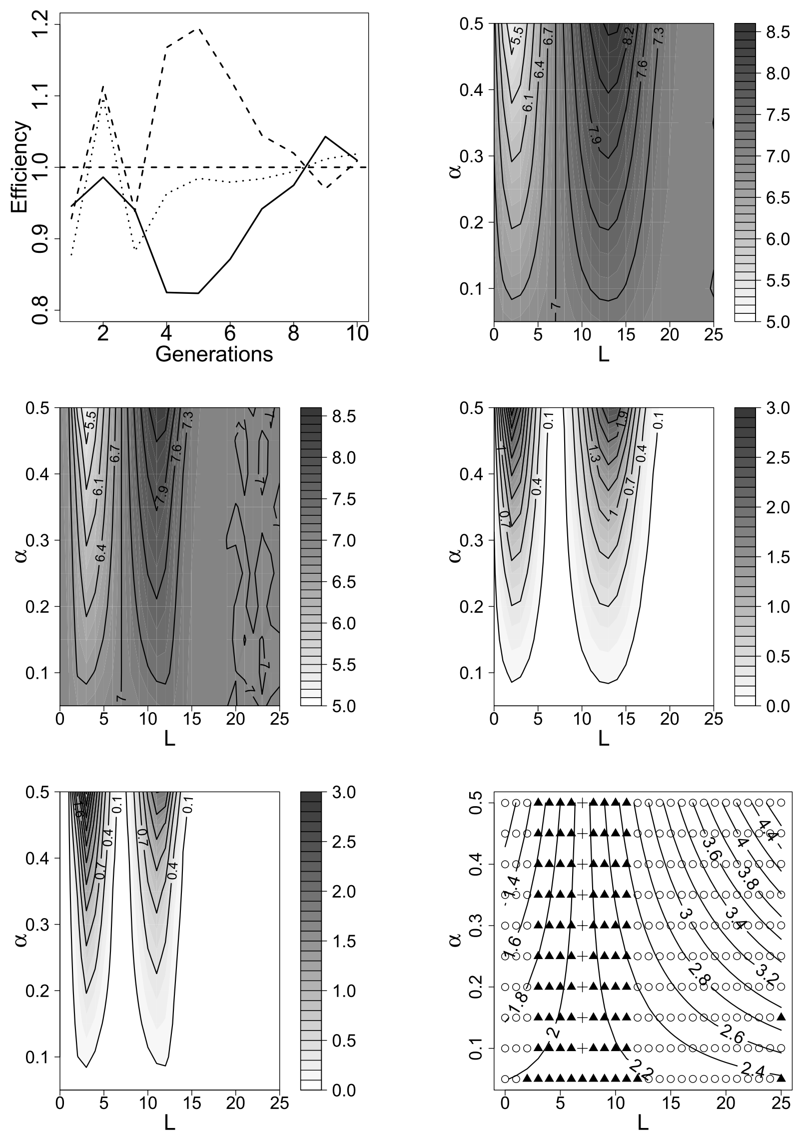

Fig 1. *First row. Left: evolution of the estimates of the relative efficiency of* $\tilde{\theta}^{NED}(\hat{p}_n)$ *to* $\tilde{\theta}^{HD}(\hat{p}_n)$ *(solid line), the relative efficiency of* $\tilde{\theta}^{HD}(\hat{p}_n)$ *to* $\tilde{\theta}^{LD}(\hat{p}_n)$ *(dashed line) and the relative efficiency of* $\tilde{\theta}^{NED}(\hat{p}_n)$ *to* $\tilde{\theta}^{LD}(\hat{p}_n)$ *(dotted line). Right: contour plot of the means of the MHDEs for each contaminated offspring distribution. Second row. Left: contour plot of the means of the MNEDEs of* $\theta_0 = 7$ *for each contaminated offspring distribution. Right: contour plot of the MSEs of the MHDEs of* $\theta_0 = 7$ *for each contaminated offspring distribution. Third row. Left: MSEs of the MNEDEs of* $\theta_0 = 7$ *for each contaminated offspring distribution. Right: Contour plot of the asymptotic mean growth rates of the contaminated models (solid line) and points* (L, α) *where the minimum of MSE of the estimates of* $\theta_0 = 7$ *by the three methods is attained in the MLDE (crosses), in the MHDE (filled triangles) and in the MNEDEs (circles).*

TABLE 1
Relative bias for Hellinger distance and negative exponential disparity for the mixture models for gross errors with $L = 0$ and different values of α.

α	$\frac{\Delta T^{HD}(\alpha,L)}{\Delta T^{LD}(\alpha,L)}$	$\frac{\Delta T^{NED}(\alpha,L)}{\Delta T^{LD}(\alpha,L)}$
−0.0001	1.0108310	0.9980851
−0.0002	1.0519480	0.9911580
−0.0003	1.1041350	0.9787859
−0.0004	1.1383230	0.9600473
−0.0005	1.1891900	0.9338905
−0.0006	1.2520860	0.9004372
−0.0007	1.3377720	0.8556661
−0.0008	1.4711780	0.7992881
−0.0009	1.7769600	0.7296631

TABLE 2
Relative bias for Hellinger distance and negative exponential disparity for the mixture models for gross errors with $L = 8$ and different values of α.

α	$\frac{\Delta T^{HD}(\alpha,L)}{\Delta T^{LD}(\alpha,L)}$	$\frac{\Delta T^{NED}(\alpha,L)}{\Delta T^{LD}(\alpha,L)}$
-0.01	1.022095	1.0004884
-0.02	1.041636	0.9985267
-0.03	1.064554	0.9947680
-0.04	1.089620	0.9895525
-0.05	1.117349	0.9830758
-0.06	1.148303	0.9737196
-0.07	1.183068	0.9622187
-0.08	1.222421	0.9480543
-0.09	1.267872	0.9306867

the models), where L is equal to 7 (consequently, the offspring mean remains unchanging). The second fact is that the most accurate method for estimating the offspring parameter does not depend on the resulting asymptotic mean growth rate after contamination, but on the state where the contamination is produced.

We have also studied the performance of these methods in presence of inliers, which correspond to the model introduced at (5.1) with $\alpha < 0$ such that $p(\theta, \alpha, L)$ is a probability distribution. To this end, we compare the potential bias, defined as $\Delta T^{\rho}(\alpha, L) = T^{\rho}(p(\theta_0, \alpha, L)) - T^{\rho}(p(\theta_0))$, with $\rho \in \{NED, HD, LD\}$. In fact, we examine the relative bias of MHDE and MNEDE with respect to MLDE under mixture model for gross errors located at $L = 0$ (Table 1), at $L = 8$ (Table 2) and at $L = 20$ (Table 3) for different values of α. The results show that the MNEDE has decreasingly less bias than the MLDE in all the cases, whereas the inliers have the opposite effect on the MHDE.

TABLE 3
Relative bias for Hellinger distance and negative exponential disparity for the mixture models for gross errors with $L = 20$ and different values of α.

α	$\frac{\Delta T^{HD}(\alpha,L)}{\Delta T^{LD}(\alpha,L)}$	$\frac{\Delta T^{NED}(\alpha,L)}{\Delta T^{LD}(\alpha,L)}$
−0.0000075	1.2087000	0.9920636
−0.0000100	1.1790380	0.9820383
−0.0000125	1.1987310	0.9682116
−0.0000150	1.2351010	0.9501191
−0.0000175	1.2755030	0.9272487
−0.0000200	1.3213920	0.8990362
−0.0000225	1.3841210	0.8648598
−0.0000250	1.4062240	0.8240357
−0.0000275	1.6524540	0.7758114

7. Concluding remarks

In the context of controlled branching processes with random control functions, assuming a general parametric framework for the offspring distribution, we have studied the minimum disparity estimation of its main parameter.

First, we have established conditions for the existence and uniqueness of MDEs for a general discrete model. Moreover, it has been established that the proposed MDEs are strongly consistent as the associated nonparametric estimators are. In particular, we have considered as the nonparametric estimator of the offspring law, the MLE based on the observation of the entire family tree until a certain generation, which is consistent under some regularity conditions. Based on this nonparametric estimator, the limiting normality of the corresponding MDEs of the offspring parameter, suitably normalized, has been also established. These results are regarded as a generalization of those given for BGW processes (see [42]), by considering the more general branching structure given by CBPs, and more disparity measures besides Hellinger one.

The MDEs proposed for the offspring parameter are appropriate robust alternatives to the MLE based on the whole family tree. Focusing our attention on the MHDE and MNEDE, through a simulated example, we show that both are robust against outliers, showing more insensitive the MNEDE to gross-errors at points far from the offspring parameter, and the MHDE when they are at points close to the same one. However, the robustness against inliers only holds for the MNEDE.

Appendix

Proof of Theorem 3.1

(i) It is immediate from the definition of $\widetilde{\Gamma}_\rho$ and the continuity of $\rho(q,\cdot)$ in C_ρ.

(ii) From $\inf_{t\in\Theta\setminus K}\rho(p(\theta^*),t) > 0$, it is deduced that $\theta^* \in K$, and hence, $\min_{t\in K}\rho(p(\theta^*),t) = 0$; consequently, $T^\rho(p(\theta^*))$ exists. Since the function $G(\cdot)$

is nonnegative and has a unique zero at 0, $\rho(p(\theta^*), \theta) = 0$ if and only if $p(\theta^*) = p(\theta)$, and from the identifiability of $\mathcal{F}_\Theta$, this can only occur when $\theta^* = \theta$.

Proof of Theorem 3.2

We present an adaptation and extension of the proofs of Proposition 2 in [1] and of Theorem 3.2 in [33], developed for general continuous models.

Let $\theta = T^\rho(q)$ (there exists and it is unique by Theorem 3.1). For each $t \in \Theta$, using the mean value theorem for the functions $h_k(y) = G(y/p_k(t) - 1)$, $y \in [0,1]$, $k \geq 0$, one can prove that $|\rho(q_n, t) - \rho(q, t)| \leq M \sum_{k=0}^{\infty} |q_{n,k} - q_k| \to 0$, as $n \to \infty$, being M an upper bound of the function $G'(\cdot)$. Hence,

$$\sup_{t\in\Theta} |\rho(q_n, t) - \rho(q, t)| \to 0, \tag{7.1}$$

obtaining that $\rho(\cdot, t)$ is continuous in l_1 for each $t \in \Theta$. From this latter, it is deduced that $q_n \in \tilde{\Gamma}_\rho$ eventually. In fact, if $q_n \notin \tilde{\Gamma}_\rho$ eventually, for all $N \in \mathbb{N}$, there exists $k_N > N$ such that

$$\inf_{t\in\Theta\setminus C_\rho} \rho(q_{k_N}, t) \leq \min_{t\in C_\rho} \rho(q_{k_N}, t),$$

therefore $q \notin \tilde{\Gamma}_\rho$, which is in contradiction with the hypotheses of the theorem. Thus, using Theorem 3.1, there exists $T^\rho(q_n)$, which we denote θ_n to ease the notation, and $\theta_n \in C_\rho$ eventually. Finally, one has to show that $\theta_n \to \theta$.

Note that if $\rho(q_n, \theta_n) \leq \rho(q, \theta)$, then $|\rho(q_n, \theta_n) - \rho(q, \theta)| = \rho(q, \theta) - \rho(q_n, \theta_n) \leq \rho(q, \theta_n) - \rho(q_n, \theta_n)$; on the other hand, if $\rho(q_n, \theta_n) \geq \rho(q, \theta)$, then $|\rho(q_n, \theta_n) - \rho(q, \theta)| = \rho(q_n, \theta_n) - \rho(q, \theta) \leq \rho(q_n, \theta) - \rho(q, \theta)$. Thus,

$$\begin{aligned} |\rho(q_n, \theta_n) - \rho(q, \theta)| &\leq |\rho(q, \theta_n) - \rho(q_n, \theta_n)| + |\rho(q_n, \theta) - \rho(q, \theta)| \\ &\leq 2 \sup_{t\in\Theta} |\rho(q_n, t) - \rho(q, t)|, \end{aligned}$$

and from (7.1), $\rho(q_n, \theta_n) \to \rho(q, \theta)$. Moreover, $|\rho(q_n, \theta_n) - \rho(q, \theta_n)| \to 0$ is also deduced from (7.1), and as a result, $\rho(q, \theta_n) \to \rho(q, \theta)$.

If the sequence $\{\theta_n\}_{n\geq 0}$ does not converge to θ, then there exists a subsequence $\{\theta_{n_j}\}_{j\in\mathbb{N}} \subseteq \{\theta_n\}_{n\in\mathbb{N}}$ such that $\theta_{n_j} \to \theta^* \neq \theta$, as $j \to \infty$. From (A1), taking into account Remark 3 *(a)*, one has that $\rho(q, \cdot)$ is continuous and $\rho(q, \theta_{n_j}) \to \rho(q, \theta^*)$, as $j \to \infty$. Due to all of the above, one has $\rho(q, \theta) = \rho(q, \theta^*)$, which contradicts the uniqueness of $T^\rho(q)$.

Proof of Theorem 3.3

It is analogous to the previous proof taking into account that (7.1) for $\rho = HD$ is followed from

$$\sup_{t\in\Theta} |HD(q_n, t)^{1/2} - HD(q, t)^{1/2}| \leq \|q_n^{1/2} - q^{1/2}\|_2.$$

Proof of Theorem 3.4

First of all, note that since $\tilde{p}_{n,k} \to p_k$ a.s., for each $k \geq 0$, by Glick's Theorem (see [10], p.10), one has $\tilde{p}_n \to p$ a.s. in l_1.

(i) It is immediate from Theorem 3.2 and the fact that $\tilde{p}_n \to p$ a.s. in l_1.

(ii) The proof is analogous to that of Theorem 3.2 in [42]. Bearing in mind Theorem 3.3, to obtain the eventual existence and the consistency it is enough to prove $||\tilde{p}_n^{1/2} - p^{1/2}||_2 \to 0$ a.s. and this is shown from the convergence of $\tilde{p}_n$ to p in l_1 and the inequality $||\tilde{p}_n^{1/2} - p^{1/2}||_2^2 \leq ||\tilde{p}_n - p||_1$.

The measurability of $\tilde{\theta}_n^{\rho}(\tilde{p}_n)$ and $\tilde{\theta}_n^{HD}(\tilde{p}_n)$ is obtained by Corollary 2.1 in [6].

Proof of Theorem 4.1 (i)

To prove *(i)* we adapt and extend the proofs of Theorem 1 in [1] and of Theorem 3.4 in [33] developed for general continuous models. In order to facilitate the proof, we will assume that $P[Z_n \to \infty] = 1$.

Let $\dot{\rho}(\hat{p}_n, \theta)$ and $\ddot{\rho}(\hat{p}_n, \theta)$ be the first and the second derivative of $\rho(\hat{p}_n, \theta)$ with respect to θ. Since $\tilde{\theta}_n^{\rho}(\hat{p}_n) = \arg\min_{\theta \in \Theta} \rho(\hat{p}_n, \theta)$, $\dot{\rho}(\hat{p}_n, \tilde{\theta}_n^{\rho}(\hat{p}_n)) = 0$, and from the Taylor series expansion of $\dot{\rho}(\hat{p}_n, \tilde{\theta}_n^{\rho}(\hat{p}_n))$ around θ_0 one obtains

$$\Delta_{n-1}^{1/2}(\tilde{\theta}_n^{\rho}(\hat{p}_n) - \theta_0) = -\Delta_{n-1}^{1/2} \dot{\rho}(\hat{p}_n, \theta_0) \ddot{\rho}(\hat{p}_n, \theta_n^*)^{-1},$$

where θ_n^* is a point between θ_0 and $\tilde{\theta}_n^{\rho}(\hat{p}_n)$. Consequently, from Slutsky's Theorem, it is enough to prove

$$\ddot{\rho}(\hat{p}_n, \theta_n^*) \xrightarrow{P} I(\theta_0), \tag{7.2}$$

$$-\Delta_{n-1}^{1/2} \dot{\rho}(\hat{p}_n, \theta_0) \xrightarrow{d} N\left(0, I(\theta_0)\right). \tag{7.3}$$

Observe that

$$\begin{aligned}
\dot{\rho}(\hat{p}_n, \theta) &= -\sum_{k=0}^{\infty} p_k'(\theta) A(\delta(\hat{p}_n, \theta, k)), \\
\ddot{\rho}(\hat{p}_n, \theta_n^*) &= -\sum_{k=0}^{\infty} p_k''(\theta_n^*) A(\delta(\hat{p}_n, \theta_n^*, k)) \\
&\quad + \sum_{k=0}^{\infty} A'(\delta(\hat{p}_n, \theta_n^*, k))(1 + \delta(\hat{p}_n, \theta_n^*, k)) u(\theta_n^*, k)^2 p_k(\theta_n^*).
\end{aligned}$$

On the one hand, from (A2) and (A3), $\delta(\hat{p}_n, \theta_n^*, k) \to 0$, and consequently, using (A5), one has $A(\delta(\hat{p}_n, \theta_n^*, k)) \to 0$ a.s. and $A'(\delta(\hat{p}_n, \theta_n^*, k)) \to 1$ a.s. for each $k \in \mathbb{N}_0$. Therefore, applying the dominated convergence theorem, (A4) and (A5), one has

$$\sum_{k=0}^{\infty} p_k''(\theta_0) A(\delta(\hat{p}_n, \theta_n^*, k)) \xrightarrow{P} 0,$$

$$\sum_{k=0}^{\infty} A'(\delta(\hat{p}_n, \theta_n^*, k))(1+\delta(\hat{p}_n, \theta_n^*, k))u(\theta_0, k)^2 p_k(\theta_0) \xrightarrow{P} \sum_{k=0}^{\infty} u(\theta_0, k)^2 p_k(\theta_0) = I(\theta_0).$$

Moreover, as θ_n^* converges to θ_0 in probability, $A(\delta)$ and $A'(\delta)(1+\delta)$ are bounded, (4.2) and (4.3),

$$\begin{aligned} \sum_{k=0}^{\infty} p_k''(\theta_n^*) A(\delta(\hat{p}_n, \theta_n^*, k)) &\xrightarrow{P} 0, \\ \sum_{k=0}^{\infty} A'(\delta(\hat{p}_n, \theta_n^*, k))(1+\delta(\hat{p}_n, \theta_n^*, k))u(\theta_n^*, k)^2 p_k(\theta_n^*) &\xrightarrow{P} I(\theta_0), \end{aligned}$$

hence, (7.2) yields.

In order to establish (7.3), since

$$\begin{aligned} -\Delta_{n-1}^{1/2} \dot{\rho}(\hat{p}_n, \theta_0) = \Delta_{n-1}^{1/2} \sum_{k=0}^{\infty} p_k'(\theta_0)\delta(\hat{p}_n, \theta_0, k) \\ + \Delta_{n-1}^{1/2} \sum_{k=0}^{\infty} p_k'(\theta_0)[A(\delta(\hat{p}_n, \theta_0, k)) - \delta(\hat{p}_n, \theta_0, k)], \end{aligned}$$

again by using Slutsky's Theorem, it is sufficient to prove that

$$\Delta_{n-1}^{1/2} \sum_{k=0}^{\infty} p_k'(\theta_0)\delta(\hat{p}_n, \theta_0, k) \xrightarrow{d} N(0, I(\theta_0)), \tag{7.4}$$

$$\Delta_{n-1}^{1/2} \sum_{k=0}^{\infty} p_k'(\theta_0)[A(\delta(\hat{p}_n, \theta_0, k)) - \delta(\hat{p}_n, \theta_0, k)] \xrightarrow{P} 0. \tag{7.5}$$

Note that due to $\sum_{k=0}^{\infty} p_k'(\theta_0) = 0$ and to (A4) (a), then

$$\Delta_{n-1}^{1/2} \sum_{k=0}^{\infty} p_k'(\theta_0)\delta(\hat{p}_n, \theta_0, k) = \Delta_{n-1}^{-1/2} \sum_{i=0}^{n-1} \sum_{j=1}^{\phi_i(Z_i)} u(\theta_0, X_{ij}).$$

Thus

$$\Delta_{n-1}^{1/2} \sum_{k=0}^{\infty} p_k'(\theta_0)\delta(\hat{p}_n, \theta_0, k) \stackrel{d}{=} \Delta_{n-1}^{-1/2} \sum_{i=0}^{\Delta_{n-1}} u(\theta_0, X_{0i}),$$

where $\stackrel{d}{=}$ indicates *equal in distribution*. Now, bearing in mind that $E[u(\theta_0, X_{0i})] = 0$, $Var[u(\theta_0, X_{0i})] = I(\theta_0)$, $\Delta_n(\tau_m^{n+1} - 1)^{-1}(\tau_m - 1) \to \tau W$ a.s., with W the limit variable introduced in (A3) (this property is deduced by applying Remark

4 (c) and Cèsaro's lemma), the hypotheses of the central limit theorem given in Theorem I in [11] hold and one has

$$\Delta_{n-1}^{-1/2} I(\theta_0)^{-1/2} \sum_{i=0}^{\Delta_{n-1}} u(\theta_0, X_{0i}) \xrightarrow{d} N(0,1).$$

Consequently, (7.4) holds.

Respect to (7.5), applying $|A(t^2-1)-(t^2-1)| \leq B(t-1)^2$ for some $B>0$ (see [30], p. 1107), one has

$$\begin{aligned}
&\left|\Delta_{n-1}^{1/2} \sum_{k=0}^{\infty} p_k'(\theta_0)[A(\delta(\hat{p}_n,\theta_0,k)) - \delta(\hat{p}_n,\theta_0,k)]\right| \\
&\qquad \leq B\Delta_{n-1}^{1/2} \sum_{k=0}^{\infty} |u(\theta_0,k)| \left(\hat{p}_{n,k}^{1/2} - p_k(\theta_0)^{1/2}\right)^2 \\
&\qquad = B\Delta_{n-1}^{-1/2} \tau_m^{n/2} \sum_{k=0}^{\infty} A_{n,k}^2, \qquad (7.6)
\end{aligned}$$

with

$$\begin{aligned}
A_{n,k}^2 &= \tau_m^{-n/2} \Delta_{n-1} |u(\theta_0,k)| \left(\hat{p}_{n,k}^{1/2} - p_k(\theta_0)^{1/2}\right)^2 \\
&= 2\tau_m^{-n/2} |s_k'(\theta_0)| \Delta_{n-1} s_k(\theta_0)^{-1} \left(\hat{p}_{n,k}^{1/2} - p_k(\theta_0)^{1/2}\right)^2, \qquad (7.7)
\end{aligned}$$

and recall $s(\theta) = p(\theta)^{1/2}$. Let us demonstrate $\sum_{k=0}^{\infty} A_{n,k}^2 =$ [1] $o_P(1)$. To this end, we prove $\lim_{n\to\infty} \sum_{k=0}^{\infty} E[A_{n,k}^2] = 0$. First, taking into account that in [18] it was proved that

$$(p_k(\theta_0)(1-p_k(\theta_0)))^{-1/2} \Delta_{n-1}^{1/2} (\widehat{p}_{n,k} - p_k(\theta_0)) \xrightarrow{d} N(0,1),$$

using delta method one obtains that $\Delta_{n-1}^{1/2}(\widehat{p}_{n,k}^{1/2} - p_k(\theta_0)^{1/2})$ converges in distribution. Thus, from $\tau_m^{-n/2} \to 0$ and Slutsky's theorem, $A_{n,k}^2 = o_P(1)$ follows.

Second, let denote $V_i(k) = \sum_{j=1}^{\phi_{i-1}(Z_{i-1})} (I_{\{X_{i-1j}=k\}} - p_k(\theta_0))$. It can be checked that $E[V_i(k)] = 0$ and $Var[V_i(k)] = p_k(\theta_0)(1-p_k(\theta_0))E[\varepsilon(Z_{i-1})]$. Moreover, using (A3)(a) and (A3)(c), for each $i \in \mathbb{N}$, $E[\varepsilon(Z_i)] \leq K_0' E[Z_i] \leq K_0 \tau_m^i$ for some constants $K_0', K_0 \in \mathbb{R}$. Consequently,

$$\begin{aligned}
E[A_{n,k}^4] &\leq 4\tau_m^{-n} |s_k'(\theta_0)|^2 p_k(\theta_0)^{-1} E\left[\left(\sum_{i=1}^{n} V_i(k)\right)^2\right] \\
&= 4|s_k'(\theta_0)|^2 (1-p_k(\theta_0)) \tau_m^{-n} E\left[\sum_{i=1}^{n} \varepsilon(Z_{i-1})\right] \\
&\leq K_0 |s_k'(\theta_0)|^2 (1-p_k(\theta_0)) \sum_{i=0}^{n-1} \tau_m^{-(n-i)}.
\end{aligned}$$

[1] We write $X_n = o_P(Y_n)$ to mean $P[|X_n| > \epsilon |Y_n|] \to 0$, as $n \to \infty$, for each $\epsilon > 0$.

Since $\sum_{i=0}^{n-1}\tau_m^{-(n-i)} = \tau_m^{-n}(\tau_m^n - 1)(\tau_m - 1)^{-1} \to (\tau_m - 1)^{-1}$, one obtains $\sup_n E[A_{n,k}^4] < \infty$ and $\{A_{n,k}\}_{n\in\mathbb{N}}$ is uniformly integrable for each $k \geq 1$. Therefore, since $A_{n,k}^2 = o_P(1)$ as $n \to \infty$, we have $A_{n,k}^2$ converges to 0 in l_1, as $n \to \infty$. Moreover, $\sum_{k=0}^{\infty} E[A_{n,k}^2] \leq K_0^{1/2}\sum_{k=0}^{\infty}|s_k'(\theta_0)| < \infty$, as a consequence, by the dominated convergence theorem, $\lim_{n\to\infty}\sum_{k=0}^{\infty} E[A_{n,k}^2] = 0$.

Now, from (7.6), since $\tau_m^{-n}\Delta_{n-1} \to (\tau_m - 1)^{-1}\tau W$ a.s., from Remark 4 (c), one has (7.5).

Proof of Theorem 4.1 (ii)

In order to prove Theorem 4.1 *(ii)*, we will make use of a previous result:

Lemma 1. *Let be a CBP satisfying (A3) and (A6), with offspring distribution $p = p(\theta_0)$. Assume (A2), $p \in \hat{\Gamma}_{HD}$ and $\theta_0 \in int(\Theta)$; then*

$$\Delta_{n-1}^{-1/2}\sum_{l=1}^{n}\sum_{j=1}^{\phi_{l-1}(Z_{l-1})} s'_{X_{l-1j}}(\theta_0)p_{X_{l-1j}}^{-1/2} \xrightarrow{d} N\left(0, ||s'(\theta_0)||_2^2\right),$$

with respect to the distribution $P[\cdot|Z_n \to \infty]$.

Proof. To simplify the proof, we assume that $P[Z_n \to \infty] = 1$. Let $\beta_l = \sum_{j=1}^{\phi_{l-1}(Z_{l-1})} s'_{X_{l-1j}}(\theta_0)p_{X_{l-1j}}^{-1/2}$ and $\mathcal{G}_l = \sigma(X_{ij}, \phi_i(k) : j \geq 1, k \geq 0, i = 0, \ldots, l-1)$; then $\{\beta_l, \mathcal{G}_l\}_{l\geq 0}$ is a martingale difference. Indeed, let $\omega \in \Omega$ such that $Z_{l-1}(\omega) = z$, one has that

$$\begin{aligned}
E[\beta_l|\mathcal{G}_{l-1}](\omega) &= E\left[\sum_{j=1}^{\phi_{l-1}(z)} s'_{X_{l-1j}}(\theta_0)p_{X_{l-1j}}^{-1/2}\right] \\
&= E\left[E\left[\sum_{j=1}^{\phi_{l-1}(z)} s'_{X_{l-1j}}(\theta_0)p_{X_{l-1j}}^{-1/2}\middle|\phi_{l-1}(z)\right]\right] \\
&= \varepsilon(z)E\left[s'_{X_{01}}(\theta_0)p_{X_{01}}^{-1/2}\right] = \varepsilon(z)\sum_{k=0}^{\infty} s_k'(\theta_0)p_k^{1/2} = 0,
\end{aligned}$$

since $\theta_0 = \arg\min_{\theta\in\Theta} HD(p, \theta)$. In addition, it satisfies

$$E[\beta_l^2|\mathcal{G}_{l-1}] = \varepsilon(Z_{l-1})||s'(\theta_0)||_2^2 \qquad \text{a.s.} \tag{7.8}$$

To show the latter, let $\omega \in \Omega$ such that $Z_{l-1}(\omega) = z$; then

$$\begin{aligned}
E[\beta_l^2|\mathcal{G}_{l-1}](\omega) &= Var[\beta_l|\mathcal{G}_{l-1}](\omega) \\
&= Var\left[\sum_{j=1}^{\phi_{l-1}(z)} s'_{X_{l-1j}}(\theta_0)p_{X_{l-1j}}^{-1/2}\right]
\end{aligned}$$

$$\begin{aligned} &= E\left[Var\left[\sum_{j=1}^{\phi_{l-1}(z)} s'_{X_{l-1j}}(\theta_0)p_{X_{l-1j}}^{-1/2}\middle|\phi_{l-1}(z)\right]\right] \\ &= E\left[\phi_{l-1}(z)Var\left[\sum_{j=0}^{\infty} s'_{X_{l-1j}}(\theta_0)p_{X_{l-1j}}^{-1/2}\right]\right] \\ &= \varepsilon(z)||s'(\theta_0)||_2^2 = \varepsilon(Z_{l-1}(\omega))||s'(\theta_0)||_2^2. \end{aligned}$$

Moreover,

$$\Delta_{n-1}^{-1/2}\sum_{l=1}^{n}\beta_l = \left(\frac{\tau_m^n}{\Delta_{n-1}}\right)^{1/2}\left[\frac{1}{\tau_m^{n/2}}\sum_{l=1}^{n}\frac{\left((\varepsilon(Z_{l-1})+1)^{1/2}-(\tau_m^{l-1}\tau W)^{1/2}\right)\beta_l}{(\varepsilon(Z_{l-1})+1)^{1/2}} + (\tau W)^{1/2}\sum_{l=1}^{n}\frac{\tau_m^{-(n-l+1)/2}\beta_l}{(\varepsilon(Z_{l-1})+1)^{1/2}}\right].$$

Because of $\tau_m^{-n}\Delta_{n-1} \to (\tau_m-1)^{-1}\tau W$ a.s., it is enough to prove

$$\sum_{l=1}^{n}\frac{\left((\varepsilon(Z_{l-1})+1)^{1/2}-(\tau_m^{l-1}\tau W)^{1/2}\right)\beta_l}{(\varepsilon(Z_{l-1})+1)^{1/2}} = o_P(\tau_m^{n/2}), \tag{7.9}$$

that is, $\tau_m^{-n/2}\sum_{l=1}^{n}\left((\varepsilon(Z_{l-1})+1)^{1/2}-(\tau_m^{l-1}\tau W)^{1/2}\right)\beta_l(\varepsilon(Z_{l-1})+1)^{-1/2}$ converges to 0 in probability, and

$$(\tau_m-1)^{1/2}\sum_{l=1}^{n}\frac{\tau_m^{-(n-l+1)/2}\beta_l}{(\varepsilon(Z_{l-1})+1)^{1/2}} \xrightarrow{d} N\left(0,||s'(\theta_0)||_2^2\right), \tag{7.10}$$

as $n\to\infty$. The proof follows similar steps to those given in Theorem 2 in [41]. For (7.9), using Cauchy-Schwarz inequality,

$$\sum_{l=1}^{n}\Big((\varepsilon(Z_{l-1})+1)^{1/2}-(\tau_m^{l-1}\tau W)^{1/2}\Big)\beta_l(\varepsilon(Z_{l-1})+1)^{-1/2} \le A_n^{1/2}B_n^{1/2},$$

with

$$\begin{aligned} A_n &= \sum_{l=1}^{n}\tau_m^{(l-1)/2}\left(((\varepsilon(Z_{l-1})+1)\tau_m^{-(l-1)})^{1/2}-(\tau W)^{1/2}\right)^2, \\ B_n &= \sum_{l=1}^{n}\tau_m^{(l-1)/2}(\varepsilon(Z_{l-1})+1)^{-1}\beta_l^2. \end{aligned}$$

On the one hand, since $(\varepsilon(Z_{l-1})+1)\tau_m^{-(l-1)})^{1/2} \to (\tau W)^{1/2}$ a.s., one has $A_n = o(\sum_{l=1}^{n}\tau_m^{(l-1)/2}) = o(\tau_m^{n/2})$.

Moreover, from (7.8), $E[B_n] = O(\tau_m^{n/2})$. As a consequence, $A_n = o(\tau_m^{n/2})$ and $B_n =$ [2]$O_P(\tau_m^{n/2})$, and hence (7.9) is proved.

To obtain (7.10), we define $\gamma_{nj} = \beta_{n-j+1}(\varepsilon(Z_{n-j})+1)^{-1/2}$, $j = 1,\dots,n$; then

$$
\begin{aligned}
(\tau_m - 1)^{1/2} \sum_{l=1}^{n} \frac{\tau_m^{-(n-l+1)/2}\beta_l}{(\varepsilon(Z_{l-1})+1)^{1/2}} &= (\tau_m - 1)^{1/2} \sum_{j=1}^{n} \tau_m^{-j/2}\gamma_{nj} \\
&= U_n^{(n)} = U_J^{(n)} + (\tau_m - 1)^{1/2} \sum_{j=J+1}^{n} \tau_m^{-j/2}\gamma_{nj},
\end{aligned}
$$

where $U_J^{(n)} = (\tau_m - 1)^{1/2} \sum_{j=1}^{J} \tau_m^{-j/2}\gamma_{nj}$, $J = 1,\dots,n$.

For $J \geq 1$ and given $(t_1,\dots,t_J) \in \mathbb{R}^J$, using analogous arguments to those given in the proof of Theorem 1 in [26], we prove that, as $n \to \infty$,

$$
E\Big[\exp\Big(i \sum_{j=1}^{J} t_j \tau_m^{-j/2}\gamma_{nj}\Big)\Big] \to \exp\Big(-\frac{1}{2}||s'(\theta_0)||_2^2 \sum_{j=1}^{J} t_j^2 \tau_m^{-j}\Big),
$$

that is, the random vector $(\tau_m^{-1/2}\gamma_{n1},\dots,\tau_m^{-J/2}\gamma_{nJ})$ converges in distribution to a J-dimensional Gaussian vector which is centered and with covariance matrix given by $||s'(\theta_0)||_2^2\ \mathrm{diag}(\tau_m^{-1},\dots,\tau_m^{-J})$; as a consequence of the application of delta method, one obtains that $U_J^{(n)} \xrightarrow{d} U_J$, with U_J following a $N(0,(\tau_m - 1)||s'(\theta_0)||_2^2 \sum_{j=1}^{J} \tau_m^{-j})$. Moreover, for every $n \geq 0$ and $\epsilon > 0$, using Chebyshev's inequality, one has

$$
\begin{aligned}
P\left[|U_J^{(n)} - U_n^{(n)}| > \epsilon\right] &\leq \epsilon^{-2}(\tau_m - 1)E\left[\left(\sum_{j=J+1}^{\infty} \tau_m^{-j/2}\gamma_{nj}\right)^2\right] \\
&= \epsilon^{-2}(\tau_m - 1)\sum_{j=J+1}^{\infty} Var\left[\tau_m^{-j/2}\gamma_{nj}\right] \\
&= \epsilon^{-2}(\tau_m - 1)||s'(\theta_0)||_2^2 \sum_{j=J+1}^{\infty} \tau_m^{-j} E\left[\frac{\varepsilon(Z_{n-j})}{\varepsilon(Z_{n-j})+1}\right] \\
&\leq \epsilon^{-2}(\tau_m - 1)||s'(\theta_0)||_2^2 \sum_{j=J+1}^{\infty} \tau_m^{-j}.
\end{aligned}
$$

As a result, there exists a constant $k_0 \geq 0$ such that

$$
\limsup_{n\to\infty} P\left[|U_J^{(n)} - U_n^{(n)}| > \epsilon\right] \leq k_0 \sum_{j=J+1}^{\infty} \tau_m^{-j} \to 0, \text{ as } J \to \infty.
$$

[2]We write $X_n = O_P(Y_n)$ to mean that for every $\epsilon > 0$ there exists a constant M such that $\sup_n P[|X_n| > M|Y_n|] < \epsilon$.

Finally, from Theorem 25.5 in [5] and the fact that $U_J \xrightarrow{d} N(0,||s'(\theta_0)||_2^2)$, as $J\to\infty$, it is verified $U_n^{(n)} \xrightarrow{d} N\left(0,||s'(\theta_0)||_2^2\right)$, as $n\to\infty$, and hence (7.10) is obtained. $\square$

Once Lemma 1 is proved, the proof of Theorem 4.1 *(ii)* is analogous to the proof of Theorem 3.4 in [42]. Recall that $T^{HD}(\hat{p}_n)=\tilde{\theta}_n^{HD}(\hat{p}_n)$ eventually exists, so applying Theorem 3.3 of [42],

$$\begin{aligned}\Delta_{n-1}^{1/2}(\tilde{\theta}_n^{HD}(\hat{p}_n) \quad &- \quad \theta_0) = \Delta_{n-1}^{1/2}\big[a_n\sum_{k=0}^{\infty}s_k'(\theta_0)(\hat{p}_{n,k}^{1/2}-p_k(\theta_0)^{1/2}) \\ &- \quad ||s'(\theta_0)||_2^{-2}\sum_{k=0}^{\infty}s_k'(\theta_0)(\hat{p}_{n,k}^{1/2}-p_k(\theta_0)^{1/2})\big]\end{aligned} \tag{7.11}$$

where $a_n\to 0$. To obtain the distribution of (7.11), one has to determine the distribution of $\sum_{k=0}^{\infty}s_k'(\theta_0)(\hat{p}_{n,k}^{1/2}-p_k(\theta_0)^{1/2})$, which verifies $\sum_{k=0}^{\infty}s_k'(\theta_0)(\hat{p}_{n,k}^{1/2}-p_k(\theta_0)^{1/2})=\sum_{k=0}^{\infty}s_k'(\theta_0)\hat{p}_{n,k}^{1/2}$ due to $\theta_0=\arg\min_{\theta\in\Theta}||p_k(\theta)^{1/2}-p_k(\theta_0)^{1/2}||_2$. From the fact that

$$\begin{aligned}\sum_{k=0}^{\infty}s_k'(\theta_0)\hat{p}_{n,k}^{1/2} \quad &= \quad \frac{1}{2}\sum_{k=0}^{\infty}s_k'(\theta_0)p_k(\theta_0)^{-1/2}\hat{p}_{n,k}^{1/2} \\ &\quad -\frac{1}{2}\sum_{k=0}^{\infty}s_k'(\theta_0)p_k(\theta_0)^{-1/2}(\hat{p}_{n,k}^{1/2}-p_k(\theta_0)^{1/2})^2,\end{aligned}$$

and $I(\theta_0)=4||s'(\theta_0)||_2^2$, it suffices

$$\Delta_{n-1}^{1/2}\sum_{k=0}^{\infty}s_k'(\theta_0)p_k(\theta_0)^{-1/2}\hat{p}_{n,k}^{1/2} \quad \xrightarrow{d} \quad N(0,||s'(\theta_0)||_2^2), \tag{7.12}$$

$$\Delta_{n-1}^{1/2}\sum_{k=0}^{\infty}s_k'(\theta_0)p_k(\theta_0)^{-1/2}(\hat{p}_{n,k}^{1/2}-p_k(\theta_0)^{1/2})^2 \quad \xrightarrow{P} \quad 0. \tag{7.13}$$

Now, from

$$\Delta_{n-1}^{1/2}\sum_{k=0}^{\infty}s_k'(\theta_0)p_k(\theta_0)^{-1/2}\hat{p}_{n,k}^{1/2}=\Delta_{n-1}^{-1/2}\sum_{l=1}^{n}\sum_{j=1}^{\phi_{l-1}(Z_{l-1})}s'_{X_{l-1j}}(\theta_0)p_{X_{l-1j}}(\theta_0)^{-1/2},$$

and Lemma 1, one has (7.12).

For each $k,n\geq 0$, we define $C_{n,k}=2^{-1}A_{n,k}$, where $A_{n,k}^2$ are the random variables introduced in (7.7). Following the same arguments as in the proof of *(i)*, one establishes (7.13) and this completes the proof.

Proof of Theorem 5.1

(i) (a) Let $\theta_L = T^\rho(p(\theta,\alpha,L))$. If the sequence $\{\theta_L\}_{L\geq 0}$ does not converge to θ, as $L\to\infty$, then there will exist a subsequence, which we continue denoting $\{\theta_L\}_{L\geq 0}$, such that $\theta_L \to \theta_1 \neq \theta$. From the definition of θ_L, for each $t\in\Theta$,

$$\rho(p(\theta,\alpha,L),\theta_L) \leq \rho(p(\theta,\alpha,L),t), \tag{7.14}$$

follows; moreover, applying a generalization of the dominated convergence theorem (see [38], p.92), one has

$$\rho(p(\theta,\alpha,L),\theta_L) \to \rho((1-\alpha)p(\theta),\theta_1), \qquad \text{as } L\to\infty; \tag{7.15}$$

as a consequence, from (7.14) and (7.15), for each $t\in\Theta$,

$$\rho((1-\alpha)p(\theta),\theta_1) \leq \rho((1-\alpha)p(\theta),t). \tag{7.16}$$

On the one hand, since $\rho^*(\alpha,p(\theta),t)=0$ if and only if $t=\theta$, $\rho^*(\alpha,p(\theta),\theta_1) > 0 = \rho^*(\alpha,p(\theta),\theta)$. On the other hand, due to $\rho((1-\alpha)p(\theta),t)$ is an increasing function of $\rho^*(\alpha,p(\theta),t)$, one obtains $\rho((1-\alpha)p(\theta),\theta_1) > \rho((1-\alpha)p(\theta),\theta)$, which contradicts (7.16).

(i) (b) The continuity of the function $L\mapsto T^\rho(p(\theta,\alpha,L))$ is immediate and the boundedness of the sequence $\{\theta_L\}_{L\geq 0}$ is deduced from its convergence.

(i) (c) By definition of θ_L, from the Taylor series expansion of $\dot\rho(p(\theta,\alpha,L),\theta_L)$ around θ one has

$$\frac{\theta_L-\theta}{\alpha} = -\frac{\alpha^{-1}\dot\rho(p(\theta,\alpha,L),\theta)}{\ddot\rho(p(\theta,\alpha,L),\theta_L^*)},$$

where θ_L^* is a point between θ and θ_L. Consequently, it will be sufficient to prove

$$\begin{aligned}\lim_{\alpha\to 0}\alpha^{-1}\dot\rho(p(\theta,\alpha,L),\theta) &= -u(\theta,L), &(7.17)\\ \lim_{\alpha\to 0}\ddot\rho(p(\theta,\alpha,L),\theta_L^*) &= I(\theta). &(7.18)\end{aligned}$$

With the same arguments as Theorem 4.1 *(i)*, one can prove

$$\begin{aligned}\dot\rho(p(\theta,\alpha,L),\theta) &= -\sum_{k=0}^\infty p_k'(\theta)A(\delta(p(\theta,\alpha,L),\theta,k)),\\ \ddot\rho(p(\theta,\alpha,L),\theta_L^*) &= -\sum_{k=0}^\infty p_k''(\theta_L^*)A(\delta(p(\theta,\alpha,L),\theta_L^*,k))\\ &\quad+\sum_{k=0}^\infty A'(\delta(p(\theta,\alpha,L),\theta_L^*,k))u(\theta_L^*,k)^2 p_k(\theta_L^*)\\ &\quad\cdot(1+\delta(p(\theta,\alpha,L),\theta_L^*,k)).\end{aligned}$$

Using L'Hôpital's rule and the fact that $\sum_{k=0}^\infty p_k'(\theta)=0$, (7.17) is obtained. To show (7.18) we use the convergence dominated theorem in the above expression.

(ii) The proof follows similar steps to Theorem 7 in [3] and it is omitted.

Acknowledgements

The authors thank the referees for their careful reading of our paper and for their constructive comments which have improved its presentation. This research has been supported by the Ministerio de Educación, Cultura y Deporte (grant FPU13/03213), Ministerio de Economía y Competividad (grant MTM2015-70522-P), the Junta de Extremadura (grant GR15105) and the FEDER.

References

[1] Basu, A., Sarkar, S. and Vidyashankar, A. N. (1997) Minimum negative exponential disparity estimation in parametric models. *Journal of Statistical Planning and Inference*, **58**, 349–370. MR1450021

[2] Basu, A. Shioya, H. and Park, C. (2011) *Statistical Inference: The Minimum Distance Approach.* Chapman & Hall/CRC. MR2830561

[3] Beran, R. (1977) Minimum Hellinger distance estimates for parametric models. *Annals of Statistics*, **5(3)**, 445–463. MR0448700

[4] Bercu, B. (1999) Weighted estimation and tracking for Bienaymè–Galton–Watson processes with adaptive control. *Statistics & Probability Letters*, **42**, 415–421. MR1707187

[5] Billingsley, P. (1979) *Probability and Measure.* Wiley and Sons, Inc. MR0534323

[6] Brown, L. D. and Purves, R. (1973) Measurable selections of extrema. *The Annals of Statistics*, **1(5)**, 902–912. MR0432846

[7] Bruss, F. and Slavtchova-Bojkova, M. (1999) On waiting times to populate an environment and a question of statistical inference. *Journal of Applied Probability*, **36**, 261–267. MR1699619

[8] Cressie, N. and Read, T. R. C. (1984) Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, **46(3)**, 440–464. MR0790631

[9] Devroye, L. (1998) *Probabilistic Methods for Algorithmic Discrete Mathematics*, volume 16 of *Algorithms and Combinatorics*, chapter Branching Processes and Their Applications in the Analysis of Tree Structures and Tree Algorithms. Springer-Verlag. MR1678582

[10] Devroye, L. and Györfi, L. (1985) *Nonparametric Density Estimation: the L_1 View.* John Willey & Sons. MR0780746

[11] Dion, J. P. (1974) Estimation of the mean and the initial probabilities of a branching process. *Journal of Applied Probability*, **11**, 687–694. MR0375488

[12] Dion, J. P. and Essebbar, B. (1995) On the statistics of controlled branching processes. *Lecture Notes in Statistics*, **99**, 14–21. MR1351256

[13] Ellis, M., Elphick, C. (2007) Using a stochastic model to examine the ecological, economic and ethical consequences of population control in a charismatic invasive species: mute swans in North America. *Journal of Applied Ecology*, **44**, 312–322.

[14] González, M., Gutiérrez, C., Martínez, R., Minuesa, C. and del Puerto, I. (2016) Bayesian analysis for controlled branching processes.

Chapter 11 in *Branching Processes and Their Applications*, volume 219, *Lecture Notes in Statistics - Proceedings*.

[15] González, M., Gutiérrez, C., Martínez, R. and del Puerto, I. (2013) Bayesian inference for controlled branching processes through MCMC and ABC methodologies. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas*, **107(2)**, 459–473. MR3199723

[16] González, M., Martínez, R. and del Puerto, I. (2004) Nonparametric estimation of the offspring distribution and mean for a controlled branching process. *Test*, **13**, 465–479. MR2154009

[17] González, M., Martínez, R. and del Puerto, I. (2005) Estimation of the variance for a controlled branching process. *Test*, **14**, 199–213. MR2203429

[18] González, M., Minuesa, C. and del Puerto, I. (2016) Maximum likelihood estimation and Expectation-Maximization algorithm for controlled branching processes. *Computational Statistics and Data Analysis*, **93**, 209–227. MR3406206

[19] González, M., Molina, M. and del Puerto, I. (2002) On the class of controlled branching processes with random control functions. *Journal of Applied Probability*, **39**, 804–815. MR1938172

[20] González, M., Molina, M. and del Puerto, I. (2003) On the geometric growth in controlled branching processes with random control function. *Journal of Applied Probability*, **40**, 995–1006. MR2012682

[21] González, M., Molina, M., and del Puerto, I. (2005) Asymptotic behaviour of critical controlled branching processes with random control functions. *Journal of Applied Probability*, **42**, 463–477. MR2145488

[22] González, M., Molina, M. and del Puerto, I. (2006) Geometric growth for stochastic difference equations with application to branching populations. *Bernoulli*, **12**, 931–942. MR2265669

[23] González, M., del Puerto, I., Martínez, R., Molina, M., Mota, M. and Ramos, A. (2010) *Workshop on Branching Processes and Their Applications*, volume 197 in *Lecture Notes in Statistics - Proceedings*. Springer-Verlag. MR2761879

[24] Guttorp, P.(1991) Statistical inference for branching processes. John Wiley and Sons, Inc. MR1254434

[25] Haccou, P., Jagers, P., and Vatutin, V. (2005) *Branching processes: variation, growth and extinction of populations.* Cambridge University Press. MR2429372

[26] Heyde, C. C. and Brown, B. M. (1971) An invariance principle and some convergence rate results for branching processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **20(4)**, 271–278. MR0310987

[27] Inamdar, A. and Kale, M. (2016) Joint estimation of offspring mean and offspring variance of controlled branching process. *Sankhya A: The Indian Journal of Statistics.* DOI 10.1007/s13171-016-0082-2. MR3575743

[28] Kale, M. M. (2000) Estimation of parameters in controlled branching process. *Stochastic Modelling and Applications*, **3**, 1–15.

[29] Kimmel, M. and Axelrod, D. E. (2015) *Branching Processes in Biology.* Springer-Verlag. MR3310028

[30] Lindsay, B. G. (1994) Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *The Annals of Statistics*, **22(2)**, 1081–1114. MR1292557

[31] Martínez, R., Mota, M. and del Puerto, I. (2009) On asymptotic posterior normality for controlled branching processes. *Statistics*, **43(4)**, 367–378. MR2547706

[32] Pardo, L. (2006) *Statistical Inference based on Divergence Measures.* Chapman & Hall/CRC, 2006. MR2183173

[33] Park, C. and Basu, A. (2004) Minimum disparity estimation: asymptotic normality and breakdown point results. *Bulletin of Informatics and Cybernetics*, **36**, 19–33. MR2139489

[34] del Puerto, I., González, M., Gutiérrez, C., Martínez, R., Minuesa, C., Molina, M., Mota, M., and Ramos, A. (2016) *Branching Processes and Their Applications*, volume 219 in *Lecture Notes in Statistics*. Springer.

[35] del Puerto, I. and Yanev, N. (2008) *Leading-Edge Applied Mathematical Modeling Research.*, chapter 11: Branching Processes with Multitype Random Control Functions: subcritical case, pages 363–374. Nova Science Publishers, Inc. MR2503807

[36] Rahimov, I. and Al-Sabah, W. (2007) Controlled branching processes with continuous states. *Journal of Applied Probability and Statistics*, **2**, 123–137. MR2439949

[37] Rao, C. (1961) Asymptotic efficiency and limiting information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 531–546. MR0133192

[38] Royden, H. L. (1988) *Real Analysis.* Macmillan Publisher Company. MR1013117

[39] Simpson, D. G. (1987) Minimum Hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association*, **82(399)**, 802–807. MR0909985

[40] Sriram, T. N. (1994) Invalidity of bootstrap for critical branching processes with immigration. *The Annals of Statistics*, **22**, 565–1114. MR1292554

[41] Sriram, T. N., Bhattacharya, A., González, M., Martínez, R. and del Puerto, I. (2007) Estimation of the offspring mean in a controlled branching process with a random control function. *Stochastic Processes and their Applications*, **117**, 928–946. MR2330726

[42] Sriram, T. N. and Vidyashankar, A. N. (2000) Minimum Hellinger distance estimation for supercritical Galton–Watson processes. *Statistics and Probability Letters*, **50**, 331–342. MR1802227

[43] Stoimenova, V., Atanasov, D. and Yanev, N. (2004) Robust estima-

tion and simulation of branching processes. *Comptes rendus de l'Acadèmie bulgare des sciences*, **57(5)**, 19–22. MR2077362

[44] Todd, C., Forsyth, D., Choquenot, D. (2007) Modelling the effect of fertility control on koala-forest dynamics. *Journal of Applied Ecology*, **45**, 568–578.

[45] Tukey, J. (1960) *Contributions to probability and statistics*, chapter A survey of sampling from contaminated distributions. Standford University Press. MR0120720

[46] Yanev, G. P. and Yanev, N. M. (1996) Branching processes with two types of emigration and state-dependent immigration. *Lecture Notes in Statistics*, **114**, 216–228. MR1466718

[47] Yanev, N. M. (1975) Conditions for degeneracy of ϕ-branching processes with random ϕ. *Theory of Probability and its Applications*, **20**, 421–428. MR0375492