

Robust Estimation in Controlled Branching Processes: Bayesian Estimators via Disparities*

Miguel González^{†,||}, Carmen Minuesa^{‡,||}, Inés del Puerto^{§,||},
and Anand N. Vidyashankar[¶]

Abstract. In this paper we describe Bayesian inferential methods for data modeled by controlled branching processes that account for model robustness via the use of disparities. Under regularity conditions, we establish that estimators obtained using disparity-based posterior, such as expected and maximum a posteriori estimates, are consistent and efficient under the posited model. Additionally, we establish that the estimates are robust to model misspecification and presence of outliers. To this end, we develop several fundamental ideas relating minimum disparity estimators to Bayesian estimators obtained using the disparity-based posterior, for dependent tree-structured data. We illustrate the methodology through a simulated example and apply our methods to a real data set from cell kinetics.

Keywords: branching process, controlled process, disparity measures, robustness, Bayesian inference.

MSC2020 subject classifications: 60J80, 62M05.

1 Introduction

Branching processes are routinely used to model population evolution in a variety of scientific disciplines such as cell biology, population demography, biochemical processes, genetics, epidemiology, and actuarial sciences (see, for instance Devroye (1998), Haccou et al. (2005), González et al. (2010), Kimmel and Axelrod (2015) and del Puerto et al. (2016)). While several variants of branching processes are available, a particularly useful variant is the controlled branching process (CBP) with random control functions, on which this work is focused.

*The authors would like to thank Dr. Jacinto Martín (University of Extremadura) for his valuable comments and suggestions.

[†]Department of Mathematics and Instituto de Computación Científica Avanzada, University of Extremadura, Avda. de Elvas s/n, 06006 Badajoz, Spain, mvelasco@unex.es

[‡]Department of Mathematics, Autonomous University of Madrid, Ciudad Universitaria de Cantoblanco, 28049 Madrid, Spain, carmen.minuesa@uam.es

[§]Department of Mathematics and Instituto de Computación Científica Avanzada, University of Extremadura, Avda. de Elvas s/n, 06006 Badajoz, Spain, idelpuerto@unex.es

[¶]Department of Statistics, Volgenau School of Engineering, George Mason University, 4400 University Drive, MS 4A7, Fairfax, VA 22030 USA, avidyash@gmu.edu

^{||}This research has been supported by the Ministerio de Educación, Cultura y Deporte (grant FPU13/03213), the Junta de Extremadura (grant GR18103 and IB16099), the Ministerio de Economía y Competitividad of Spain (grant MTM2015-70522-P), the Ministerio de Ciencia e Innovación of Spain (grant PID2019-108211GB-I00/AEI/10.13039/501100011033) and the Fondo Europeo de Desarrollo Regional.

These processes are discrete time and discrete state stochastic processes – like classical Galton-Watson processes (GWPs) – describing the growth of a population over the generations using probability distributions for the reproduction. However, unlike the GWPs, the number of progenitors is determined by a random mechanism which is referred to as control functions. This fact allows for modelling real-life phenomenon such as random migratory movements (for instance, immigration and emigration), with great flexibility. Moreover, although the GWP and the CBP have a Markovian structure and individuals reproduce independently of the others in both models, in a CBP the evolution of the tree generated by a progenitor is not independent of the trees generated by the others. This lack of independence leads to substantial technical challenges in the theoretical developments and is described in detail in the manuscript. A good reference for the probabilistic theory and inferential issues developed until now for the CBPs is the recent monograph González et al. (2018). The great flexibility offered by CBPs comes with a cost; they require specification of multiple distributions such as the offspring distribution and the control distributions. Divergence-based methods have been used to provide methodologies for inference in these settings that are robust to presence of outliers and efficient when the posited model is correct, see Sriram and Vidyashankar (2000) for the GWP and the more general approach given in González et al. (2017) for the CBP. Analysing robust procedures in depth in the field of branching processes is of great interest since it is not unusual to find the existence of a small proportion of individuals in a population whose reproductive capacity is influenced by temporary events, leading to outliers in the data. While the aforementioned papers deal with the problem from the frequentist standpoint, divergence-based methods in the Bayesian framework in the context of branching processes with strong theoretical guarantees do not exist. This paper is focused on addressing this problem.

Bayesian inference in the presence of model misspecification has received much attention in recent years. In the context of classification problems, Jiang and Tanner (2008) studied the behaviour of the so-called Gibbs posterior under a variety of conditions on the risk functions and hence allowing for potential model misspecification. See also the classical reference Catoni (2004) and Bissiri et al. (2017). On the other hand, Hooker and Vidyashankar (2014) provided an alternative approach for Bayesian inference under model misspecification using divergences for conditionally independent and identically distributed (c.i.i.d.) random variables. Indeed, in their work they evaluate the effect of model misspecifications by studying the asymptotic behaviour of the posterior estimates under the posited model and misspecified model settings. Alternatively, again in the c.i.i.d. setting, Ghosh and Basu (2016) and Ghosh and Basu (2017) used power divergences to derive robust Bayesian methods. More recently, Miller and Dunson (2018) developed an alternative coarsening approach, for c.i.i.d. random variables, which essentially amounts to assuming that the observed data are within an ϵ -neighbourhood of posited model for a suitably defined neighbourhood and ϵ .

Returning to the setting of CBPs, in order to make inference of the offspring distribution, we assume that it belongs to some known general parametric family. This is a reasonable assumption in the context of branching processes, since we often have some knowledge on the reproduction process of the population. For instance, the geometric distribution is proposed as a reproduction law for a GWP applied to model data from a yeast cell colony in Guttorp (1991), or is used to establish analogies between

branching processes and the thermodynamic properties of small systems in Corral et al. (2016). Additionally, a GWP with a generalized power series offspring distribution is considered as a model for the spread of an infectious disease in a population in Angelov and Slavtchova-Bojkova (2012). The classical reference of Holgate and Lakhani (1967) provides a broad description of parametric families of offspring distributions suitable in an ecological context. For example, a Poisson distribution is the natural offspring distribution to consider in a population where each progenitor produces a large number of eggs whose probability of becoming an individual is small and it is constant for all the eggs; a Bernoulli distribution with support $\{0, 2\}$ is appropriate to model the reproduction of organisms by binary fission, such as bacteria, amoeba, etc. Additional examples involving a compound Poisson distribution, a negative binomial distribution or a zero-modified geometric distribution can be found in Holgate and Lakhani (1967).

The current paper describes the use of disparities for CBP data to provide robust estimation of the parameters of the offspring distribution. We assume that the number of parents giving birth to exactly k offspring is known for all $k \geq 0$. For sake of brevity, henceforth this sample will be referred as the entire family tree. Roughly speaking, our proposed method, analogous to the c.i.i.d. case, consists in replacing the log-likelihood in the expression for the posterior distribution with an appropriately scaled disparity measure, and studying the expectation and the mode of the resulting disparity based posteriors to obtain the Bayes estimators, known as EDAP and MDAP estimators, respectively (see Section 3 for definition). However, unlike the c.i.i.d. case, two new issues arise: the scaling for the disparity is random and its asymptotic behavior depends on the parameters of interest. To address these two issues, it is common to use conditioning arguments; however, as explained previously, the lack of independence between the trees generated by the progenitor and other members of the population (due to the introduction of control functions) causes sufficient challenges and requires development of new mathematical techniques. Indeed, to establish the asymptotic properties, we need results that invoke the branching structure – see Lemma 5 in Supplementary Material (González et al., 2020) for details – and results that establish L^1 convergence of estimated posterior densities to the Gaussian density (see Theorem 4.2). It is worth mentioning that such results are not known even for likelihood based posteriors and to the best of our knowledge, the results in this paper are the first ones for dependent tree-structured data. As a special case, when the control function yields the population size of a generation, one obtains results for classical GWPs.

Moreover, turning to the assumptions, it is also important to mention that our class of probability distributions, Γ^* , defined below in Section 3, refines and improves the condition (A5) in Hooker and Vidyashankar (2014). This is related to the identifiability of the model.

Beyond the challenges outlined above, we re-emphasize that even though the asymptotic properties of Bayes estimators are similar to those of the frequentist estimators, the proofs concerning Bayes estimators are highly non-trivial. Indeed, one of the biggest challenges is to verify that the posterior mean is close to the minimum disparity estimator at a specified rate (see Theorem 4.3 below). Furthermore working with the posterior mode introduces additional difficulties and requires uniform convergence of posterior densities, which we establish in this paper. Indeed, under further conditions we show

that this yields the closeness of the posterior mode and the disparity estimator at the required rate (see Theorems 4.4 and 4.5 below).

Our algorithm to obtain the EDAP and MDAP estimators is quite different from those developed in Hooker and Vidysankar (2014). Unlike their work, which relies on Metropolis-Hastings algorithm, we use numerical techniques such as Newton-Cotes methods and non-linear optimization methods to obtain the estimators. This approach decreases the computational burden significantly especially when replacing the likelihood by alternative disparities. Furthermore, we establish that the proposed methods yield estimators that are similar in numerical value to the traditional Bayesian estimators in the absence of data contaminations, and are more robust in the presence of outliers. More details are provided in Section 6.

Besides this Introduction, the rest of the paper is structured as follows: Section 2 describes the CBP model and states the hypotheses considered throughout the manuscript. Section 3 is concerned with the description of Bayes estimators under disparity measures and provides sharp probabilistic bounds for differences between the EDAP and MDAP estimators and the frequentist minimum disparity estimators. Section 4 is devoted to the asymptotic properties of the estimators while Section 5 deals with robustness properties. Section 6 contains two examples where we use the R programming environment. Our first example is concerned with a real data set of a oligodendrocyte cell population, firstly used in Yakovlev et al. (2008). This example illustrates the versatility of the proposed methodology since such data are typically modeled by a multi-type branching process. The second example is a simulation study to illustrate the robustness of the proposed methodology. Finally, the contributions of the manuscript are summarized in Section 7. The Appendix gathering some results on the existence and continuity of functionals related to EDAP and MDAP estimators, additional figures, simulated data, the proofs of the main results and the computational codes developed for the examples are presented in the Supplementary Material.

2 The probability model

A *controlled branching process (CBP) with random control functions* is a Markov chain $\{Z_n\}_{n \in \mathbb{N}_0}$ defined recursively as follows:

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

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $N \in \mathbb{N}$, $\{X_{nj} : n \in \mathbb{N}_0; j \in \mathbb{N}\}$ and $\{\phi_n(k) : n, k \in \mathbb{N}_0\}$ are two independent families of non-negative integer valued random variables defined on a probability space (Ω, \mathcal{A}, P) . The empty sum in (1) is taken to be 0. The random variables X_{nj} , $n \in \mathbb{N}_0$, $j \in \mathbb{N}$, are assumed to be i.i.d. and $\{\phi_n(k)\}_{k \in \mathbb{N}_0}$, $n \in \mathbb{N}_0$, are independent stochastic processes with the same one-dimensional probability distributions. Intuitively, this process models an evolving population in which each individual reproduces, independently of each other and of the previous generation population, according to the same probability distribution. However, unlike the classical GWP, the number of reproducing individuals in the n -th generation, $\phi_n(Z_n)$, is a random function

$\phi_n(\cdot)$ of the generation size Z_n rather than Z_n itself. In the case when $\phi_n(x) \equiv x$, one obtains the classical GWP. As in the GWP, the quantity X_{nj} can be interpreted as the number of offspring produced by the j -th progenitor in the n -th generation. The terminology, number of progenitors of the n -th generation will sometimes be used to describe $\phi_n(Z_n)$. We denote the offspring distribution by $p = \{p_k\}_{k \in \mathbb{N}_0}$, $p_k = P[X_{01} = k]$, $k \in \mathbb{N}_0$. Moreover, in relation to the moments of the process, we denote by $m = E[X_{01}]$ and $\sigma^2 = Var[X_{01}]$, the offspring mean and variance (assumed to be finite), while by $\varepsilon(k) = E[\phi_0(k)]$ and $\sigma^2(k) = Var[\phi_0(k)]$, $k \in \mathbb{N}_0$, the mean and variance function of the random control functions.

As explained in the introduction, we focus on offspring distributions that are parametric; that is, we assume that $p_k(\theta) = p_k = P[X_{01} = k]$, for $k \in \mathbb{N}_0$, $\theta \in \Theta$ and $\Theta \subseteq \mathbb{R}$ with a non-empty interior. We notice here that extension to multidimensional case is possible at the cost of more cumbersome notations. We denote the parametric family by $\mathcal{F}_\Theta = \{\mathbf{p}_\theta : \theta \in \Theta\}$, where $\mathbf{p}_\theta = \{p_k(\theta) : k \in \mathbb{N}_0\}$ is the offspring distribution for each $\theta \in \Theta$, and these satisfy regularity conditions such as differentiability in θ . We also assume that the parametric model satisfies the following identifiability condition:

$$p_k(\theta_1) = p_k(\theta_2), \quad \forall k \in \mathbb{N}_0 \quad \Rightarrow \quad \theta_1 = \theta_2. \tag{2}$$

The main results that we describe in the paper will require the generation sizes to diverge to infinity with positive probability. While the assumption of supercriticality of the offspring distribution is sufficient in the GWP case, one needs a slight modification of such a condition (see Chapter 3 in González et al. (2018)). Assumption 1 in Section 4 fixes the framework we need in relation to this issue.

The following additional notation regarding the sample is needed for the theoretical developments:

$$\mathcal{Z}_n^* = \{Z_l(k) : 0 \leq l \leq n - 1; k \in \mathbb{N}_0\}, \quad \text{with} \quad Z_l(k) = \sum_{i=1}^{\phi_l(Z_l)} I_{\{X_{li}=k\}},$$

where I_A represents the indicator function of the set A . Notice that $Z_l(k)$ represents the number of individuals in generation l who have exactly k offspring. Also, let

$$Y_l(k) = \sum_{j=0}^l Z_j(k), \quad Y_l = \sum_{j=0}^l Z_j, \quad \text{and} \quad \Delta_l = \sum_{j=0}^l \phi_j(Z_j), \quad l, k \in \mathbb{N}_0.$$

We notice that $Y_l(k)$ represents the total number of progenitors who have exactly k offspring up to generation l . Furthermore, Y_l and Δ_l represent the total number of individuals and the total number of progenitors until the l -th generation.

3 Bayesian estimators using disparity measures

In this section, we describe Bayesian estimators using disparities. To this end, let \mathcal{L}^1 denote the space of measurable functions which are Lebesgue integrable and equipped with the L^1 -norm. Also, let $\pi(\cdot)$ denote a prior density on Θ which belongs to \mathcal{L}^1 , i.e., $\int_\Theta |\theta| \pi(\theta) d\theta < \infty$. Then, using the Bayes Theorem and Markov property it can be seen

that the posterior density of θ , given the sample \mathcal{Z}_n^* is,

$$\pi(\theta|\mathcal{Z}_n^*) \propto e^{-\Delta_{n-1}KL(\hat{p}_n, \theta)}\pi(\theta), \quad (3)$$

where $KL(q, \theta)$ is the Kullback-Leibler divergence between a probability distribution q defined on \mathbb{N}_0 and \mathbf{p}_θ ; that is,

$$KL(q, \theta) = \sum_{k=0}^{\infty} \log\left(\frac{q_k}{p_k(\theta)}\right) q_k,$$

and $\hat{p}_n = \{\hat{p}_{k,n}\}_{k \in \mathbb{N}_0}$ is the non-parametric maximum likelihood estimator (MLE) of the offspring distribution, p , based on the sample \mathcal{Z}_n^* (see González et al. (2016)); that is,

$$\hat{p}_{k,n} = \frac{Y_{n-1}(k)}{\Delta_{n-1}}, \quad k \in \mathbb{N}_0, \quad n \in \mathbb{N}. \quad (4)$$

Two traditional summary Bayes estimators for the parameter θ are (i) the posterior mean or expectation a posteriori (EAP) and (ii) the posterior mode or maximum a posteriori (MAP). These are defined as follows:

$$\theta_n^* = \int_{\Theta} \theta \pi(\theta|\mathcal{Z}_n^*) d\theta, \quad \text{and} \quad \theta_n^+ = \arg \max_{\theta \in \Theta} \pi(\theta|\mathcal{Z}_n^*), \quad n \in \mathbb{N},$$

respectively. However, it is widely known, that these estimators fail to yield robust estimates in the i.i.d. case. We illustrate this fact for the offspring distribution of the CBP in the following example.

Example 1. Consider a CBP starting with $Z_0 = 1$ individual and for each $k \in \mathbb{N}_0$ the control variable $\phi_n(k)$ follows Poisson distribution with parameter λk , with $\lambda = 0.3$. In practice, these control functions are suitable to describe an environment with an expected emigration. For the offspring distribution, we consider a mixture model for gross errors (see (12) in Section 5 for definition); specifically, each individual at each generation produces offspring according to a geometric distribution with parameter $\theta_0 = 0.3$ with probability 0.85, and with probability 0.15 it gives birth to $L = 11$ offspring. The offspring mean and variance, ignoring the presence of the contamination at the point $L = 11$, are $m = m(\theta_0) = (1 - \theta_0)/\theta_0 = 2.333$ and $\sigma^2 = \sigma^2(\theta_0) = (1 - \theta_0)/\theta_0^2 = 7.778$, respectively.

Using the statistical software R, we have simulated the first 45 generations of such a model, z_{45}^* . The evolution of the number of individuals and progenitors is shown in Figure 1 (left), where growth in both groups is observed. Next, in order to estimate the posterior density function of θ using the sample z_{45}^* , we used a beta prior distribution with parameters 1/2 and 1/2. This choice was motivated by the fact that, according to Berger and Bernardo (1992), this is a non-informative prior distribution for a parameter between 0 and 1. We do not discuss further about the choice of the prior distribution, but a sensitivity analysis is provided in the Supplementary Material to show the negligible impact of the prior distribution. The estimate of the posterior density function of θ is plotted in Figure 1 (centre), where one notices the poor estimate for the true offspring

parameter, θ_0 . For a better illustration, in Figure 1 (right) we have also plotted the evolution of the EAP and MAP estimates for θ_0 through the generations. We remark that, similar to the case for the MLE in a frequentist setting, EAP and MAP estimators also possess unsatisfactory behavior in the presence of contamination. However, these estimators possess desirable asymptotic properties – consistency and asymptotic normality – in a contamination-free context.

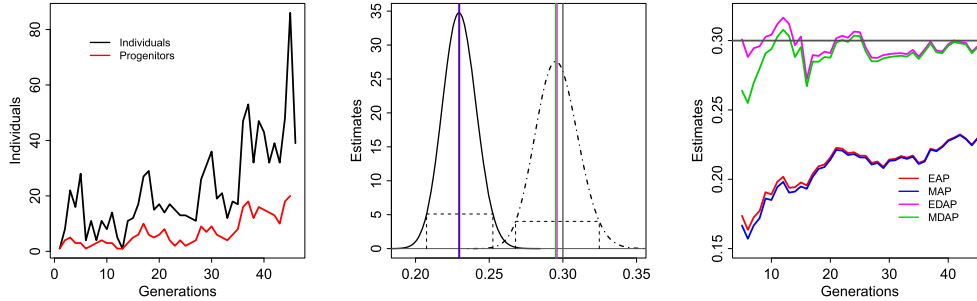


Figure 1: Left: evolution of the number of individuals (black line) and progenitors (red line). Centre: estimate of the posterior density (black line) and D -posterior density (dashed-dotted line) – see details in Subsection 6.2 – of θ given the sample z_{45}^* , together with EAP and MAP estimates (red and blue lines, resp.) and D -posterior estimates – EDAP and MDAP (magenta and green lines, resp.) – with high posterior density (HPD) intervals (dashed lines) and true value of θ_0 (dark gray line). Right: evolution of EAP, MAP, EDAP and MDAP estimates (red, blue, magenta and green lines, resp.) of θ_0 through the generations, where the horizontal dark gray line represents the true value of the parameter.

Figure 1 illustrates the undesirable properties of EAP and MAP (see red and blue lines, respectively) as well as the good ones of the D -posterior estimators (called EDAP and MDAP, see magenta and green lines, respectively) introduced and studied in Sections below. In Subsection 6.2 we also describe the probability distributions and the resulting behavior of D -posterior, not only for the contaminated models but also when there is no contamination and the contamination levels increase. These are illustrated in Figure 7 and 8 of the Section 5.4 in the Supplementary Material. We notice that the D -posterior smoothly down weighs the effect of outlying observation and as the contamination reduces the credible interval gets closer to the one determined by the classical posterior. For another aspect of this example see Subsection 6.2 below.

Situations similar to the above example clearly showcase the need of robust estimators against outlier contamination. Since outlier contamination can frequently be described using a mixture model, the robustness of the Bayes estimators to outliers can be cast in the general framework of model misspecification. The general disparity-based approach, as described in Hooker and Vidyashankar (2014) and adapted to the family tree data \mathcal{Z}_n^* generated by the CBP, facilitates such an analysis. As in Hooker and Vidyashankar (2014), the equation (3) suggests defining a density function by replacing

the Kullback-Leibler divergence with a suitable disparity measure D defined on $\Gamma \times \Theta$, where Γ is the set of all probability distributions defined on the non-negative integers. A *disparity measure* is defined (see Lindsay (1994)) using a strictly convex and thrice differentiable function $G : [-1, \infty) \rightarrow \mathbb{R}$ satisfying $G(0) = 0$ as follows:

$$D : \Gamma \times \Theta \rightarrow [0, \infty],$$

$$(q, \theta) \mapsto D(q, \theta) = \sum_{k=0}^{\infty} G(\delta(q, \theta, k)) p_k(\theta),$$

where $q = \{q_k\}_{k \in \mathbb{N}_0}$, and $\delta(q, \theta, k)$ denotes the ‘‘Pearson residual at k ’’, that is,

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

The resulting function after replacing the Kullback-Leibler divergence with a disparity-based quantity is referred to as *D-posterior density function at \hat{p}_n* and is defined as:

$$\pi_D^n(\theta|\hat{p}_n) = \frac{e^{-\Delta_{n-1} D(\hat{p}_n, \theta)} \pi(\theta)}{\int_{\Theta} e^{-\Delta_{n-1} D(\hat{p}_n, \theta)} \pi(\theta) d\theta}. \tag{5}$$

More generally, the previous definition can be extended for an arbitrary probability distribution $q \in \Gamma$, yielding the *D-posterior density function at q* that is given by

$$\pi_D^n(\theta|q) = \frac{e^{-\Delta_{n-1} D(q, \theta)} \pi(\theta)}{\int_{\Theta} e^{-\Delta_{n-1} D(q, \theta)} \pi(\theta) d\theta}. \tag{6}$$

Observe that (6) is well defined for each $q \in \Gamma$ if and only if $\text{supp}(\pi) \cap \{\theta \in \Theta : D(q, \theta) < \infty\}$ has non-null Lebesgue measure, where $\text{supp}(\pi)$ denotes the support of the prior density function. In particular, if the disparity measure D is bounded on $\Gamma \times \Theta$, then $\text{supp}(\pi) \cap \{\theta \in \Theta : D(q, \theta) < \infty\} = \text{supp}(\pi)$, for each $q \in \Gamma$, and the density function in (6) is well defined. Taking into consideration this fact, we focus our attention on the following set:

$$\tilde{\Gamma} = \{q \in \Gamma : \text{supp}(\pi) \cap \{\theta \in \Theta : D(q, \theta) < \infty\} \text{ has non-null Lebesgue measure}\}.$$

Analogous to the classical Bayes estimators, one can summarize the information from the *D-posterior* using the estimators *expectation a D-posteriori (EDAP)* and *maximum a D-posteriori (MDAP)*; these are defined as follows:

- *Expectation a D-posteriori (EDAP)*:

$$\theta_n^{*D} = \int_{\Theta} \theta \pi_D^n(\theta|\hat{p}_n) d\theta, \quad n \in \mathbb{N}. \tag{7}$$

- *Maximum a D-posteriori (MDAP)*:

$$\theta_n^{+D} = \arg \max_{\theta \in \Theta} \pi_D^n(\theta|\hat{p}_n), \quad n \in \mathbb{N}. \tag{8}$$

Specifically, the EDAP estimator mimics the Bayes point estimators for the posterior density under the squared error loss function, whereas the MDAP estimator does it under 0 – 1 loss function. We next focus on a few useful examples of non-negative disparity measures and the corresponding D -posterior density functions and refer to Cressie and Read (1984) and Lindsay (1994) for additional examples.

Example 2. (a) *The Kullback-Leibler divergence with the parametric family \mathcal{F}_Θ , defined above, is provided by the function $G(\delta) = (\delta + 1) \log(\delta + 1) - \delta$. Consequently, the D -posterior distribution and the EDAP and MDAP estimators using this disparity coincide with the posterior distribution and the EAP and MAP estimators, respectively.*

(b) *The disparity determined by the function $G(\delta) = 2[(\delta + 1)^{1/2} - 1]^2$ is the twice squared Hellinger distance (or simply Hellinger distance), denoted by $HD(q, \theta)$, namely $HD(q, \theta) = 2 \sum_{k=0}^\infty (q_k^{1/2} - p_k^{1/2}(\theta))^2$, for each $q \in \Gamma$, and $\theta \in \Theta$. The D -posterior distribution using this disparity is referred to as HD -posterior distribution and the EDAP and MDAP estimators are denoted by θ_n^{*HD} and θ_n^{+HD} , respectively, for each $n \in \mathbb{N}$.*

(c) *The disparity defined by using the function $G(\delta) = e^{-\delta} - 1 + \delta$ is known as negative exponential disparity and denoted by $NED(q, \theta)$ for each $q \in \Gamma$, and $\theta \in \Theta$. The D -posterior distribution using this disparity is referred to as NED -posterior distribution and the EDAP and MDAP estimators are denoted by θ_n^{*NED} and θ_n^{+NED} , respectively, for each $n \in \mathbb{N}$.*

For the study of asymptotic and robustness properties of the EDAP and MDAP estimators, which we address in Sections 4 and 5, we introduce the *EDAP functions*, defined for each $n \in \mathbb{N}$ as follows:

$$\begin{aligned} \bar{T}_n : \Gamma \times \Omega &\rightarrow \Theta, \\ (q, \omega) &\mapsto \bar{T}_n(q)(\omega) = \frac{\int_{\Theta} \theta e^{-\Delta_{n-1}(\omega)D(q,\theta)} \pi(\theta) d\theta}{\int_{\Theta} e^{-\Delta_{n-1}(\omega)D(q,\theta)} \pi(\theta) d\theta}, \end{aligned}$$

and the *MDAP functions*, defined as:

$$\begin{aligned} \tilde{T}_n : \Gamma \times \Omega &\rightarrow \Theta, \\ (q, \omega) &\mapsto \tilde{T}_n(q)(\omega) = \arg \min_{\theta \in \Theta} (\Delta_{n-1}(\omega)D(q, \theta) - \log(\pi(\theta))), \end{aligned}$$

whenever this minimum exists. With these definitions, $\theta_n^{*D}(\omega) = \bar{T}_n(\hat{p}_n)(\omega)$, and $\theta_n^{+D}(\omega) = \tilde{T}_n(\hat{p}_n)(\omega)$. Note that these functions depend on the total number of progenitors, and one has different EDAP and MDAP functions for each disparity measure and for each prior distribution; however, in order to ease the notation we will not make explicit this relation.

Despite the similarities between EDAP and MDAP functions defined in the context of branching processes and those in the case of i.i.d. random variables, there is a key difference. While in the i.i.d. case these are deterministic functions, in the current

context they are random variables. This randomness arises due to the random sample size Δ_{n-1} and leads to questions concerning the existence and measurability of these functionals that must be treated carefully (see, for instance Brown and Purves (1973)). However, in order to ease the reading of the paper we omit those details here and we relegate the study of those questions to the Appendix in the Supplementary Material. Taking into account these considerations, we assume the conditions that guarantee the existence and uniqueness of $\tilde{T}_n(q)$, $q \in \Gamma$, throughout the paper.

4 Consistency and asymptotic normality

In this section, we focus our attention on the asymptotic behaviour of the EDAP and MDAP estimators. The proofs of the results of this section are presented in Section 2 of the Supplementary Material. One of our main results in this section is that EDAP and MDAP estimators are asymptotically efficient when the posited offspring distribution belongs to the parametric family. This will be obtained as a consequence of a more general result; *viz.*, the \mathcal{L}^1 almost sure convergence of posterior density to a Gaussian density with mean 0 and variance equal to the inverse of the Fisher information.

As stated in Section 2, we impose the following assumption on the asymptotic behaviour of the process:

Assumption 1 (On the supercriticality of CBPs). *The CBP satisfies:*

- (a) *There exists $\tau = \lim_{k \rightarrow \infty} \varepsilon(k)k^{-1} < \infty$ and the sequence $\{\sigma^2(k)k^{-1}\}_{k \in \mathbb{N}}$ is bounded.*
- (b) *$\tau_m = \tau m > 1$ and Z_0 is large enough so that $P[Z_n \rightarrow \infty] > 0$.*
- (c) *$\{Z_n \tau_m^{-n}\}_{n \in \mathbb{N}}$ converges a.s. to a finite random variable W such that $P[W > 0] > 0$.*
- (d) *$\{W > 0\} = \{Z_n \rightarrow \infty\}$ a.s.*

In González et al. (2016), it is established that under the Assumption 1, $\Delta_n \rightarrow \infty$ a.s., and $\hat{p}_{k,n}$ is a strongly consistent estimator of p_k , for each $k \in \mathbb{N}_0$, on the set $\{Z_n \rightarrow \infty\}$. Recall that if the offspring distribution generating the CBP is parametric, then there exists an interior point $\theta_0 \in \Theta$ such that $p = \mathbf{p}_{\theta_0}$.

To establish the results, we need additional regularity conditions and hence, in the remainder of this paper we assume that for each $q \in \Gamma$, the first and the second derivative of $D(q, \theta)$ with respect to θ exist and we denote them by $\dot{D}(q, \theta)$ and $\ddot{D}(q, \theta)$; the reader is referred to Section 4 in González et al. (2017) for conditions that guarantee their existence. We also denote by $I^D(\theta) = \ddot{D}(p, \theta)$, and $I_n^D(\theta) = \ddot{D}(\hat{p}_n, \theta)$, where p is the posited offspring distribution and \hat{p}_n was defined in (4). Henceforth, we also assume – without loss of generality (see Lindsay (1994), p. 1089) – that $G'(0) = 0$ and $G''(0) = 1$. Thus, taking into account the previous hypotheses and the identifiability condition (2), if $p = \mathbf{p}_{\theta_0}$, one has that $I^D(\theta_0)$ reduces to the Fisher information at θ_0 denoted by $I(\theta_0)$ (see González et al. (2017)).

In the proofs of the results in this section the asymptotic properties of the frequentist minimum disparity estimator will play a critical role. As it was introduced in González et al. (2017), the *minimum disparity estimator (MDE)* of θ_0 based on \hat{p}_n is defined as:

$$\hat{\theta}_n^D = \arg \min_{\theta \in \Theta} D(\hat{p}_n, \theta),$$

and the associated *disparity function* defined as:

$$\begin{aligned} T : \Gamma &\rightarrow \Theta \\ q &\mapsto T(q) = \arg \min_{\theta \in \Theta} D(q, \theta), \end{aligned}$$

whenever this minimum exists. For details about conditions for the existence, uniqueness, and continuity of the function T , we refer the reader to Theorems 3.1–3.3 in González et al. (2017). Note that analogous to EDAP and MDAP functions, the disparity function depends on the disparity measure which we suppress in our notations. With this notation, $T(\hat{p}_n) = \hat{\theta}_n^D$.

Furthermore, we strengthen our regularity conditions on the disparity measure and the prior distribution as below:

Assumption 2 (On the disparity function and the prior distribution).

- (a) For $q \in \Gamma$, $T(q)$ exists, is unique and $T(q) \in \text{int}(\Theta)$ (interior of Θ).
- (b) $\pi(\cdot)$ is bounded.
- (c) $\pi(\cdot)$ is thrice differentiable, the third derivative of $\pi(\cdot)$ is bounded, and $\pi(T(q)) > 0$.

Assumption 3 (On the continuity of the disparity measure and the MDE). *The following properties concerning the disparity measure and the MDE hold:*

- (a) There exists some $C > 0$ such that for each $q_1, q_2 \in \Gamma$, $\sup_{\theta \in \Theta} |D(q_1, \theta) - D(q_2, \theta)| \leq C \|q_1 - q_2\|_1$.
- (b) Let $\theta_p = T(p)$ satisfy $I^D(\theta_p) > 0$ and $I_n^D(\theta) \rightarrow I^D(\theta_p)$ a.s. on $\{Z_n \rightarrow \infty\}$, as $\theta \rightarrow \theta_p$ and $n \rightarrow \infty$.
- (c) The MDE satisfies that

$$\hat{\theta}_n^D \rightarrow \theta_p \text{ a.s. on } \{Z_n \rightarrow \infty\}, \tag{9}$$

$$\Delta_{n-1}^{1/2}(\hat{\theta}_n^D - \theta_p) \xrightarrow{d} N(0, I^D(\theta_p)^{-1}), \tag{10}$$

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

Remark 4.1. (a) We note that (a) in Assumption 3 holds for Hellinger distance or any disparity, defined by $G(\cdot)$ satisfying $G(\cdot)$ and $G'(\cdot)$ are bounded in $[-1, \infty)$ (see González et al. (2017), p. 313).

- (b) Sufficient conditions for (b) in Assumption 3 can be found in Section 4 in González et al. (2017).
- (c) Conditions that guarantee the convergence of (9) and (10) were established in Theorems 3.4 and 4.1, respectively, in González et al. (2017).

We now introduce the following subset of probability distributions which we need in the statement of our next theorem. Let

$$\Gamma^* = \left\{ q \in \tilde{\Gamma} : \forall \eta > 0, \exists \rho > 0 \text{ such that } \inf_{|\theta - T(q)| > \eta} (D(q, \theta) - D(q, T(q))) > \rho \right\}. \quad (11)$$

In the following $\varphi(t; \theta)$ denotes the density function of a normal distribution with mean 0 and variance $I^D(\theta)^{-1}$. Additionally, let $\varphi_n(t)$ denote the density function of a normal distribution with mean 0 and variance $I_n^D(\hat{\theta}_n^D)^{-1}$. The following results hold true for the offspring distribution p without assuming that it belongs to the parametric family. In the case when $p = \mathbf{p}_{\theta_0}$ and under the identifiability condition (2), then the same results hold with $\theta_p = \theta_0$ and $I^D(\theta_p) = I(\theta_0)$. We state our first main result of this section.

Theorem 4.2. Let $\bar{\pi}_D^n(\cdot | \hat{p}_n)$ denote the D -posterior density function of $t = \Delta_{n-1}^{1/2}(\theta - \hat{\theta}_n^D)$. Let $p \in \Gamma^*$ satisfy (a) in Assumption 2. If Assumptions 1 and 3 hold, then:

- (i) $\int |\bar{\pi}_D^n(t | \hat{p}_n) - \varphi(t; \theta_p)| dt \rightarrow 0$ a.s. on $\{Z_n \rightarrow \infty\}$.
- (ii) $\int t |\bar{\pi}_D^n(t | \hat{p}_n) - \varphi(t; \theta_p)| dt \rightarrow 0$ a.s. on $\{Z_n \rightarrow \infty\}$.
- (iii) $\int |\bar{\pi}_D^n(t | \hat{p}_n) - \varphi_n(t)| dt \rightarrow 0$ a.s. on $\{Z_n \rightarrow \infty\}$.

The following theorem shows that the EDAP estimator mimics the MDE at the rate $\Delta_{n-1}^{1/2}$. This feature leads to the asymptotic normality of the centered and scaled EDAP estimator.

Theorem 4.3. Under the hypotheses of Theorem 4.2, the following convergences hold:

- (i) $\Delta_{n-1}^{1/2}(\theta_n^{*D} - \hat{\theta}_n^D) \rightarrow 0$ a.s. on $\{Z_n \rightarrow \infty\}$.
- (ii) $\Delta_{n-1}^{1/2}(\theta_n^{*D} - \theta_p) \xrightarrow{d} N(0, I^D(\theta_p)^{-1})$ with respect to $P[\cdot | Z_n \rightarrow \infty]$.

To establish the asymptotic properties of the MDAP estimator, we need the uniform convergence of posterior density to the Gaussian density. This is the content of our next theorem.

Theorem 4.4. Under conditions of Theorem 4.2, if (b) and (c) in Assumption 2 hold, then:

$$\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |\bar{\pi}_D^n(t | \hat{p}_n) - \varphi(t; \theta_p)| = 0 \quad \text{a.s. on } \{Z_n \rightarrow \infty\}.$$

Now, we turn to the asymptotic behaviour of MDAP estimators.

Theorem 4.5. *Under the hypotheses of Theorem 4.4 and if Θ is complete and separable, then the following convergences hold:*

- (i) $\Delta_{n-1}^{1/2}(\theta_n^{+D} - \hat{\theta}_n^D) \rightarrow 0$ a.s. on $\{Z_n \rightarrow \infty\}$.
- (ii) $\Delta_{n-1}^{1/2}(\theta_n^{+D} - \theta_p) \xrightarrow{d} N(0, I^D(\theta_p)^{-1})$ with respect to $P[\cdot | Z_n \rightarrow \infty]$.

5 Robustness properties

In this section we describe the robustness properties of Bayesian disparity estimators EDAP and MDAP. The proofs of the results of this section are gathered in Section 3 of Supplementary Material. In the context of i.i.d. data some of these issues were investigated in Hooker and Vidyashankar (2014). The study of disparities in the frequentist context for branching processes has been investigated for the Hellinger distance in Sri-ram and Vidyashankar (2000) and for a general standpoint in González et al. (2017). It is pertinent to note here that the standard robustness concepts such as influence function, breakdown point (both finite and asymptotic) and α -influence curves, $\alpha \in (0, 1)$, take a familiar form as the i.i.d. case and hence we will keep the discussion rather brief.

We begin with a brief description of the framework. As is common and widely accepted in the study of robustness problems, we focus on the gross error contamination model. This model is a particular case of the semiparametric gross error introduced in Hampel (1974), and is given by

$$p(\theta, \alpha, L) = (1 - \alpha)\mathbf{p}_\theta + \alpha\eta_L, \tag{12}$$

where $\theta \in \Theta$, $\alpha \in (0, 1)$, $L \in \mathbb{N}_0$, and η_L is a point mass distribution at L . A key ingredient used in the investigation of robustness is the α -influence function, where $0 < \alpha < 1$. The α -influence function of a random variable $\bar{T} : \Gamma \times \Omega \rightarrow \Theta$, is a mapping such that

$$IF_\alpha(\cdot, \bar{T}, p) : \mathbb{N}_0 \times \Omega \rightarrow \mathbb{R},$$

$$(L, \omega) \mapsto IF_\alpha(L, \bar{T}, p)(\omega) = \frac{\bar{T}(p(\theta_0, \alpha, L))(\omega) - \bar{T}(\mathbf{p}_{\theta_0})(\omega)}{\alpha}.$$

In our first result in this section we consider the disparity function T as a degenerate random variable, i.e., $T(q)(\omega) = T(q)$, for each $\omega \in \Omega$ and each $q \in \Gamma$, and establish that the α -influence function of the EDAP (respectively, MDAP) function and of the disparity function are close and characterize the difference in terms of the random sample size Δ_{n-1} . To this end, we need the following additional assumption.

Assumption 4 (On the differentiability of the disparity measure). *The disparity measure satisfies:*

- (a) For each $q \in \Gamma$, $\ddot{D}(q, \cdot)$ is continuous in Θ , and $\ddot{D}(q, T(q)) > 0$.
- (b) D is a disparity measure associated with a function $G(\cdot)$ which satisfies that $G(\cdot)$ and $G'(\cdot)$ are bounded on $[-1, \infty)$.
- (c) The function $\delta \in [-1, \infty) \mapsto (\delta + 1)G'(\delta) - G(\delta)$ is bounded.

The following result is an immediate consequence of Theorem 3 in the Appendix in the Supplementary Material.

Proposition 5.1. *Suppose that Assumption 1 and (c) in Assumption 2 hold. Let us fix $\alpha \in (0, 1)$ and $L \in \mathbb{N}_0$ and assume that $p(\theta_0, \alpha, L) \in \Gamma^*$, and that it satisfies (a) in Assumption 2 and (a) in Assumption 4. The following hold as $n \rightarrow \infty$ on $\{Z_n \rightarrow \infty\}$.*

- (i) $|IF_\alpha(L, \bar{T}_n, p) - IF_\alpha(L, T, p)| = o(\Delta_{n-1}^{-1/2})$ a.s.
- (ii) If Θ is complete and separable and (b) in Assumption 2 holds, then, $|IF_\alpha(L, \tilde{T}_n, p) - IF_\alpha(L, T, p)| = o(\Delta_{n-1}^{-1/2})$ a.s.

An immediate consequence of this proposition is that for each $\alpha \in (0, 1)$, and $n \in \mathbb{N}$ large enough, the α -influence curve of the disparity functional at p provides a good approximation of the α -influence curves of \bar{T}_n and of \tilde{T}_n at this probability distribution. In addition, under conditions given in Theorem 5.1 in González et al. (2017), one has that $\lim_{L \rightarrow \infty} \lim_{n \rightarrow \infty} IF_\alpha(L, \bar{T}_n, p) = 0$ a.s. and $\lim_{L \rightarrow \infty} \lim_{n \rightarrow \infty} IF_\alpha(L, \tilde{T}_n, p) = 0$ a.s.

The study of the α -influence curves as a measure for the robustness of an estimator was driven by the fact that the influence functions do not always provide a good description of the robustness properties of an estimator. Nonetheless, we establish a sufficient condition for the influence functions of the EDAP estimators to be bounded, and consequently, from the classical viewpoint, the EDAP estimators are robust. Recall the definition of the influence function for EDAP estimators at p is given by

$$IF(\cdot, \bar{T}_n, p) : \mathbb{N}_0 \times \Omega \rightarrow \mathbb{R},$$

$$(L, \omega) \mapsto IF(L, \bar{T}_n, p)(\omega) = \lim_{\alpha \rightarrow 0} IF_\alpha(L, \bar{T}_n, p)(\omega).$$

Theorem 5.2. *If (b) in Assumption 4 holds, then $|IF(L, \bar{T}_n, p)| < \infty$ a.s., for each $L \in \mathbb{N}_0$ and $n \in \mathbb{N}$.*

Next, we turn our attention to the study of breakdown point. While the notion of influence curves represents the effect of a single outlier in the asymptotic behaviour of the estimator, the breakdown point intuitively represents the percentage of contamination that the estimator can asymptotically bear without taking arbitrarily large values. Classically, the breakdown point of a general function \bar{T} at $q \in \Gamma$ is defined as:

$$B(\bar{T}, q) = \sup\{\alpha \in (0, 1) : b(\alpha, \bar{T}, q) < \infty\},$$

where $b(\alpha, \bar{T}, q) = \sup\{|\bar{T}((1 - \alpha)q + \alpha\bar{q}) - \bar{T}(q)| : \bar{q} \in \Gamma\}$. Note that $b(\alpha, \bar{T}, q) = \infty$ is equivalent to the existence of a sequence of probability distributions $\{\bar{q}_L\}_{L \in \mathbb{N}_0}$ satisfying

$|\overline{T}((1 - \alpha)q + \alpha\overline{q}_L) - \overline{T}(q)| \rightarrow \infty$, as $L \rightarrow \infty$. That sequence is called sequence of contaminating probability distributions. The notion of breakdown point can be extended in a natural way to a function $\overline{T} : \Gamma \times \Omega \rightarrow \Theta$, such that for each $q \in \Gamma$, $\overline{T}(q)$ is a random variable. We will study the breakdown point of the EDAP and MDAP functions when the contaminating probability distributions and the parametric family satisfy the following assumptions:

Assumption 5 (On the parametric family and the sequence of contaminating probability distributions).

- (a) $\lim_{L \rightarrow \infty} \sum_{k=0}^{\infty} \min(\overline{q}_{L,k}, p_k) = 0$.
- (b) $\lim_{L \rightarrow \infty} \sup_{|\theta| \leq c} \sum_{k=0}^{\infty} \min(\overline{q}_{L,k}, p_k(\theta)) = 0, \quad \forall c > 0$.
- (c) $\lim_{|\theta| \rightarrow \infty} \sum_{k=0}^{\infty} \min(p_k, p_k(\theta)) = 0$.

We notice here that the above conditions represent the worst possible contamination context (see, for instance, Park and Basu (2004), p. 28). The next result, which in content is analogous to the i.i.d. case, describes the role of the prior in the robustness of the EDAP and MDAP estimators. It is to be noted that the asymptotic breakdown point of the minimum disparity estimators is at least 50%.

Theorem 5.3. *Suppose that (c) in Assumption 4 holds true.*

- (i) *If $D(q, \theta)$ is bounded for any $q \in \Gamma$ and $\theta \in \Theta$, then the breakdown point of the EDAP at p is 1.*
- (ii) *Assume that (b) in Assumption 4, and (c) in Assumption 5 hold. Then, for any family of contaminating distributions $\{\overline{q}_L\}_{L \in \mathbb{N}_0}$ satisfying (a) and (b) in Assumption 5, the breakdown point of the MDAP at p is 1. The result also holds for the Hellinger distance.*

We next turn our attention to a new notion for the study of robustness for a sequence of estimators which are dependent on the sample size. The *asymptotic breakdown point* for a sequence of estimators $\{T_n\}_{n \in \mathbb{N}}$ at $q \in \Gamma$ is defined as:

$$B(\{T_n\}_{n \in \mathbb{N}}, q) = \sup\{\alpha \in (0, 1) : \limsup_{n \rightarrow \infty} b(\alpha, T_n, q) < \infty\}.$$

Before we describe the behaviour of EDAP and MDAP with respect to the above measure, we introduce the following assumption:

Assumption 6 (On the sequence of contaminating probability distributions). *The sequence of contaminating probability distributions $\{\overline{q}_L\}_{L \in \mathbb{N}_0}$ satisfies that for any $\alpha \in (0, 1)$ there exists $\delta > 0$ such that*

$$\inf_{L \in \mathbb{N}_0} \inf_{\theta \in \Theta} D(\alpha\overline{q}_L, \theta) > \inf_{\theta \in \Theta} D((1 - \alpha)p, \theta) + \delta.$$

Our next theorem describes the asymptotic breakdown point of EDAP and MDAP estimators. To that end, we need some additional regularity conditions on the parametric family; indeed, assume condition (A4) – alternatively, (A6) for the Hellinger distance – in González et al. (2017).

Theorem 5.4. *Suppose that Assumption 1, (c) in Assumption 2, (b) and (c) in Assumption 4, and (c) in Assumption 5 hold. Let $q \in \Gamma^*$ satisfy (a) in Assumption 4. Assume further that any family of contaminating distributions $\{\bar{q}_L\}_{L \in \mathbb{N}_0}$ satisfies (a) and (b) in Assumption 5, and Assumption 6. Then the following hold:*

- (i) *The asymptotic breakdown point of $\{\bar{T}_n\}_{n \in \mathbb{N}}$ at p is equal to the breakdown point of T at p .*
- (ii) *If Θ is complete and separable and (b) in Assumption 2 holds, then the asymptotic breakdown point of $\{\tilde{T}_n\}_{n \in \mathbb{N}}$ at p is equal to the breakdown point of T at p .*

The result also holds for the Hellinger distance.

While the above results are concerned with EDAP and MDAP estimators, in the context of the methods proposed in the manuscript it is important to know the asymptotic behaviour of the corresponding D -posterior density functions at a contaminating model. Our next result addresses this issue. Specifically, given $\alpha \in (0, 1)$, and a family of contaminating distributions $\{\bar{q}_L\}_{L \in \mathbb{N}_0}$, the following result establishes the asymptotic behaviour, as $L \rightarrow \infty$, of the D -posterior density function at $(1 - \alpha)p + \alpha\bar{q}_L$.

Theorem 5.5. *Let $\alpha \in (0, 1)$. Suppose that Assumption 1, (b) and (c) in Assumption 4, and (c) in Assumption 5 hold. Let $q \in \Gamma^*$ satisfy (a) in Assumption 2 and (a) in Assumption 4. Additionally, assume that the family of contaminating distributions $\{\bar{q}_L\}_{L \in \mathbb{N}_0}$ satisfies (a) and (b) in Assumption 5, and Assumption 6. Then the following hold:*

- (i) $\lim_{L \rightarrow \infty} \pi_D^n(\theta|(1 - \alpha)p + \alpha\bar{q}_L) = \pi_D^n(\theta|(1 - \alpha)p)$.
- (ii) $\lim_{L \rightarrow \infty} \int_{\Theta} |\pi_D^n(\theta|(1 - \alpha)p + \alpha\bar{q}_L) - \pi_D^n(\theta|(1 - \alpha)p)| d\theta = 0$.

The result also holds for the Hellinger distance.

Example 1 revisited: To illustrate the results of Theorem 5.5, in the data analysis example described in Section 6 below, we will revisit Example 1. We consider a slight modification of the model described previously, namely the offspring distribution is a mixture of $(1 - \alpha) \cdot 100\%$ of a geometric distribution of parameter 0.3 and $\alpha \cdot 100\%$ contamination at L , and the control distributions are Poisson of parameter $\lambda k = 0.45k$, where α is fixed at 0.01 and L increases. Concretely, in Figure 8 in Subsection 5.4 in Supplementary Material, note that values of L near $m = 2.333$ are not extreme and are the plausible values of a geometric distribution of parameter 0.3. However as L increases, the EAP and MAP based on D -posterior are unaffected while that based on the ordinary posterior *breakdown*.

6 Examples

In this section, we illustrate the methodology proposed via two examples. The first example is based on a real data set in the context of cell kinetics populations. In this example, we show that the EDAP and MDAP estimators based on different disparity measures are as good as the ones based on the Kullback-Leibler divergence in absence of contamination. The second example of this section, which is a continuation of Example 1 introduced in Section 3, demonstrates the accuracy of the method in a contamination context. Thus, unlike the EAP and MAP estimators based on the likelihood disparity, the twofold goodness (in contamination and free-contamination contexts) of EDAP and MDAP estimators is exemplified. In both cases, the results are illustrated by considering the Hellinger distance and the negative exponential disparity. The proposed methodology was implemented in R programming environment. The corresponding codes can be found in the Supplementary Material.

6.1 Example – Oligodendrocytes

In this example, we consider data from a study on cell proliferation published by Hyrien et al. (2006) and re-analyzed in Yakovlev et al. (2008). Both the papers were concerned with modeling the proliferation of oligodendrocyte precursor cells and their transformation into terminally differentiated oligodendrocytes through multitype age-dependent branching processes. The cell populations considered consist of two types of cells: the oligodendrocyte precursor cells, referred as type T_1 cells, and the terminally differentiated oligodendrocytes, referred as type T_2 cells. The development of both types of cells is as follows: type T_1 cells can die without any offspring or divide under normal conditions; when stimulating to division, precursor cells are capable of producing either direct progeny (two daughter cells of the same type) or a single terminally differentiated nondividing oligodendrocyte. The data are supplied by time-lapse video recording of oligodendrocyte populations. These experimental techniques enable us to record all observable events (division, differentiation, or death), as well as their timing, in the development of each individual cell. Due to the design of these experimental techniques it is reasonable to assume absence of contamination in the observed sample (the multi-daughter cell division can be observed, although rarely, in cancer studies – see Tse et al. (2012) – this is not the case in our observed sample). A special feature is the presence of censoring effects due to migration of precursor cells out of the microscopic field of observation, modelled as a process of emigration of the type T_1 cells.

Yakovlev et al. (2008) proposed frequentist estimators for the offspring distribution of the continuous time evolution by embedding it in a discrete-time multitype branching process. They incorporated cell-emigration in an artificial way by modifying the reproduction law. As a more natural alternative, we propose a two-type controlled branching process to describe the embedded discrete branching structure of the continuous time branching process previously described. In the framework given by this process, we address the problem of estimating the offspring distribution of the cell population from a Bayesian standpoint by making use of disparity measures.

Specifically, we consider a two-type controlled branching process with binomial control, which is a particular case of the model introduced in González et al. (2005). The process is defined recursively as follows:

$$Z_0 = (N, 0), \quad Z_{n+1} = \sum_{j=1}^{\phi_n(Z_n)} (X_{n,j}^{(1)}, X_{n,j}^{(2)}), \quad n \in \mathbb{N}_0,$$

where $Z_n = (Z_n^{(1)}, Z_n^{(2)})$, $N \in \mathbb{N}$, and $\{(X_{n,j}^{(1)}, X_{n,j}^{(2)}) : j \in \mathbb{N}, n \in \mathbb{N}_0\}$ and $\{\phi_n(z) : n \in \mathbb{N}_0, z \in \mathbb{N}_0^2\}$ are two independent families of non-negative integer valued random variables satisfying the following conditions:

- (i) For each $z = (z^{(1)}, z^{(2)}) \in \mathbb{N}_0^2$, the random variables $\{\phi_n(z) : n \in \mathbb{N}_0\}$ are i.i.d. following a binomial distribution with parameters $z^{(1)}$ and $\gamma \in (0, 1)$.
- (ii) The random vectors $\{(X_{n,j}^{(1)}, X_{n,j}^{(2)}) : j \in \mathbb{N}\}$, $n \in \mathbb{N}_0$ are i.i.d. whose probability distribution is as follows:

$$\begin{aligned} p_0 &= P \left[X_{n,j}^{(1)} = 0, X_{n,j}^{(2)} = 0 \right], \\ p_1 &= P \left[X_{n,j}^{(1)} = 2, X_{n,j}^{(2)} = 0 \right], \\ p_2 &= P \left[X_{n,j}^{(1)} = 0, X_{n,j}^{(2)} = 1 \right]. \end{aligned}$$

- (iii) If $n_1, n_2 \in \mathbb{N}_0$ are such that $n_1 \neq n_2$, then, the sequences $\{(X_{n_1,j}^{(1)}, X_{n_1,j}^{(2)}) : j \in \mathbb{N}\}$ and $\{(X_{n_2,j}^{(1)}, X_{n_2,j}^{(2)}) : j \in \mathbb{N}\}$ are independent.

Intuitively, $Z_n^{(j)}$ denotes the number of cells of type T_j , $j = 1, 2$, in the n th generation, and $X_{n,j}^{(1)}$ and $X_{n,j}^{(2)}$ represent the number of cells of type T_1 and T_2 , respectively, produced by the j th progenitor of type T_1 in the generation n . The random variables $\phi_n(z)$ are introduced to model the cell emigration, and therefore to determine the number of progenitors of type T_1 cells in the n th generation provided the population size at that generation is the vector z . Consequently, γ is the probability that a cell of type T_1 completes successfully its mitotic-cycle regardless of its outcome (and then $1 - \gamma$ is the probability of emigration of that T_1 cell).

We focus our attention on the offspring distribution, $\{p_0, p_1, p_2\}$, where p_0 is the probability that a cell of type T_1 dies, p_1 is the probability that a cell of type T_1 divides into two cells of type T_1 and p_2 is the probability that a cell of type T_1 differentiates into a new cell of type T_2 . Using the time-lapse video recording, one can observe the entire family tree for that process, that is, the following sample $Z_n^* = \{Z_l^{(1)}(0), Z_l^{(1)}(2), \Lambda_l : l = 0, \dots, n - 1\}$, where

$$Z_l^{(1)}(j) = \sum_{i=1}^{\phi_l(Z_l)} I_{\{X_{l,i}^{(1)}=j, X_{l,i}^{(2)}=0\}}, \quad j = 0, 2; \quad \Lambda_l = \sum_{i=1}^{\phi_l(Z_l)} I_{\{X_{l,i}^{(1)}=0, X_{l,i}^{(2)}=1\}}.$$

Note that $\phi_l(Z_l) = Z_l^{(1)}(0) + Z_l^{(1)}(2) + \Lambda_l$, $Z_l^{(1)}(j)$ is the number of cells of type T_1 in the l -th generation having exactly j offspring of type T_1 , $j = 0, 2$, and Λ_l is the number of cells of type T_1 in the l th generation having exactly one offspring of type T_2 . Our aim is to estimate the offspring distribution by using disparity measures, to provide HPD regions for the corresponding D -posterior densities and to estimate the probability that the production of precursor cells follows a supercritical probability distribution.

Using the Markov property and the independence between the emigration and reproduction phases, one can see that the likelihood function $f(\mathcal{Z}_n^*|p_0, p_1, \gamma)$ satisfies

$$f(\mathcal{Z}_n^*|p_0, p_1, \gamma) \propto p_0^{Y_{n-1}^{(1)}(0)} p_1^{Y_{n-1}^{(1)}(2)} (1 - p_0 - p_1)^{\Psi_{n-1}} \gamma^{\Delta_{n-1}} (1 - \gamma)^{Y_{n-1}^{(1)} - \Delta_{n-1}},$$

where $Y_{n-1}^{(1)}(j) = \sum_{l=0}^{n-1} Z_l^{(1)}(j)$ is the total number of cells of type T_1 in the first n generations having exactly j offspring of type T_1 , $j = 0, 2$, $\Psi_{n-1} = \sum_{l=0}^{n-1} \Lambda_l$ is the total number of cells of type T_1 in the first n generations having exactly one offspring of type T_2 , $\Delta_{n-1} = \sum_{l=0}^{n-1} \phi_l(Z_l)$ is the total number of observed progenitor cells of type T_1 in the first n generations and $Y_{n-1}^{(1)} = \sum_{l=0}^{n-1} Z_l^{(1)}$ is the total number of individuals of type T_1 in the first n generations. Moreover, observe that $\Delta_{n-1} = Y_{n-1}^{(1)}(0) + Y_{n-1}^{(1)}(2) + \Psi_{n-1}$. Thus, the MLEs of p_0, p_1, p_2 and γ based on the sample \mathcal{Z}_n^* are given, respectively, by

$$\hat{p}_{0,n} = \frac{Y_{n-1}^{(1)}(0)}{\Delta_{n-1}}, \quad \hat{p}_{1,n} = \frac{Y_{n-1}^{(1)}(2)}{\Delta_{n-1}}, \quad \hat{p}_{2,n} = \frac{\Psi_{n-1}}{\Delta_{n-1}}, \quad \hat{\gamma} = \frac{\Delta_{n-1}}{Y_{n-1}^{(1)}}.$$

Let us write $p = \{p_0, p_1, p_2\}$ and $\hat{p}_n = \{\hat{p}_{0,n}, \hat{p}_{1,n}, \hat{p}_{2,n}\}$. Consider $q = \{q_0, q_1, q_2\}$ a probability distribution on the state space $\{(0, 0), (2, 0), (0, 1)\}$, $\pi(\cdot)$ a prior distribution on the space of the probability distributions defined on such a state space and $\theta = (q_0, q_1)$. With analogous arguments as those in Section 3, we define the D -posterior density function of $\theta = (q_0, q_1)$ as

$$\pi_D^n(\theta|\hat{p}_n) = \frac{e^{-\Delta_{n-1}D(\hat{p}_n, \theta)} \pi(\theta)}{\int_{\Theta} e^{-\Delta_{n-1}D(\hat{p}_n, \theta)} \pi(\theta) d\theta}, \tag{13}$$

where $\int_{\Theta} e^{-\Delta_{n-1}D(\hat{p}_n, \theta)} \pi(\theta) d\theta = \int_{\Theta} e^{-\Delta_{n-1}D(\hat{p}_n, q)} \pi(q) dq_0 dq_1$ and $\Theta = \{(x_0, x_1) \in (0, 1) \times (0, 1) : x_0 + x_1 \leq 1\}$. Note that in this case, the parametric family which we consider is $\mathcal{F}_{\Theta} = \{\{q_0, q_1, 1 - q_0 - q_1\} : (q_0, q_1) \in \Theta\}$. As a consequence, we propose as EDAP estimators of $\theta_0 = (p_0, p_1)$ the following ones:

$$(p_{0,n}^{*D}, p_{1,n}^{*D}) = \left(\int_0^1 q_0 \pi_D^n(q_0|\hat{p}_n) dq_0, \int_0^1 q_1 \pi_D^n(q_1|\hat{p}_n) dq_1 \right), \tag{14}$$

and as MDAP estimators:

$$(p_{0,n}^{+D}, p_{1,n}^{+D}) = \arg \max_{\theta \in \Theta} \pi_D^n(\theta|\hat{p}_n),$$

where $\pi_D^n(\theta|\hat{p}_n)$ was introduced in (13), and $\pi_D^n(q_0|\hat{p}_n)$ and $\pi_D^n(q_1|\hat{p}_n)$ are the marginal density functions of q_0 and q_1 , respectively.

The data that we consider are provided from two experiments developed under two different experimental conditions: in the first one, the cells were cultured in a vehicle solution without agent stimulating differentiation, whereas in the second experiment, they were cultured with a specific hormone promoting oligodendrocyte generation. The first experiment started with 34 precursor cells cultured in a vehicle solution (with no agent promoting oligodendrocyte generation) and whose evolution was observed until the generation $n = 7$. On the other hand, the initial number of cells of type T_1 in experiment 2 was 30, which were cultured in solution with a specific hormone promoting oligodendrocyte generation and whose branching tree was observed until the generation $n = 5$. Based on the observed branching trees, we computed the non-parametric MLE for the parameters related to the reproduction and the emigration processes, which are given in Table 2.

	n	N	$Y_{n-1}^{(1)}$	Δ_{n-1}	$Y_{n-1}^{(1)}(0)$	$Y_{n-1}^{(1)}(2)$	Ψ_{n-1}
Experiment 1	7	34	425	410	158	201	51
Experiment 2	5	30	276	269	37	133	99

Table 1: Observed branching tree for each experiment.

	n	$\hat{p}_{0,n}$	$\hat{p}_{1,n}$	$\hat{p}_{2,n}$	$\hat{\gamma}_n$
Experiment 1	7	0.3854	0.4902	0.1244	0.9647
Experiment 2	5	0.1375	0.4944	0.3680	0.9746

Table 2: MLE for the main parameters of the model.

Using the MLE of p , $\hat{p}_n = \{\hat{p}_{0,n}, \hat{p}_{1,n}, \hat{p}_{2,n}\}$, for each experiment, we estimated the D -posterior density functions of (q_0, q_1) at \hat{p}_n for the Hellinger distance, the negative exponential disparity and the Kullback-Leibler divergence. Since the D -posterior density function is known, we generated 10^3 – uniformly distributed – points from the set Θ and determined the value of the D -posterior density functions at them. In all the cases, we consider the Dirichlet distribution with parameter $(1/2, 1/2, 1/2)$ as the prior distribution due to the fact that no prior knowledge about the offspring distribution – apart from its support – is available (see Berger and Bernardo (1992)). Moreover, in order to estimate the integral in (13) we used the Monte Carlo integration method, by drawing a sample of $2 \cdot 10^6$ points from the previous prior distribution. To that end, we used the function `rdirichlet()` from the package `gtools` (see Warnes et al. (2018)).

The contour plots of these D -posterior density functions for experiments 1 and 2 are represented in Figures 1 and 2 respectively, of Subsection 4.1 in the Supplementary Material, along with the 95% HPD regions for (q_0, q_1) . We also calculated the EDAP and MDAP estimators of the parameters p_0 and p_1 for the aforementioned disparity measures, which are presented in Tables 3 and 4 for the first and the second experiment, respectively. To compute the EDAP estimates, we used again the Monte Carlo integration method based on a sample of $2 \cdot 10^6$ points, whereas to determine the MDAP estimates we applied numerical methods, and more specifically, we used the function `nloptr()` from the package `nloptr` (see Ypma et al. (2018) for details). Note that, the

EDAP and MDAP estimates are close to the non-parametric estimators in Table 2. As was explained at the beginning of example, it is reasonable to assume that the observed data has no contamination. Under this assumption, one would expect, as proved in Theorems 4.2–4.5, the estimators based on KL posterior, HD posterior, and NED posterior to be similar. A numerical illustration of this issue via a simulated example is presented in the next subsection. Therefore, in a contamination-free context, it seems that the choice of the disparity measure has no impact in the estimates obtained, as expected.

Finally, we performed a sensitivity analysis to examine the influence of the parameters of Dirichlet distribution on EDAP and MDAP estimates. We mention here that there were no significant changes to the estimates due to changes in the parameters. These results and additional details are provided in Subsection 4.2 of the Supplementary Material.

Parameter	HD		NED		KL	
	EDAP	MDAP	EDAP	MDAP	EDAP	MDAP
p_0	0.3851	0.3854	0.3848	0.3854	0.3850	0.3854
p_1	0.4893	0.4902	0.4893	0.4902	0.4891	0.4902
p_2	0.1256	0.1244	0.1258	0.1244	0.1259	0.1244

Table 3: Estimates of p_0 , p_1 and p_2 based on the EDAP and MDAP estimators for HD, NED and KL in the first experiment.

Parameter	HD		NED		KL	
	EDAP	MDAP	EDAP	MDAP	EDAP	MDAP
p_0	0.1391	0.1375	0.1397	0.1375	0.1399	0.1375
p_1	0.4929	0.4944	0.4924	0.4944	0.4925	0.4944
p_2	0.3680	0.3680	0.3679	0.3680	0.3676	0.3680

Table 4: Estimates of p_0 , p_1 and p_2 based on the EDAP and MDAP estimators for HD, NED and KL in the second experiment.

An interesting issue to tackle in this Bayesian framework is to determine the probability of having a supercritical offspring distribution governing the production of precursor cells. This is an important problem since if the mean of the variable $X_{0,1}^{(1)}$, $m = 2p_1$, is less than 1, then the population of cells of type T_1 will become extinct with probability 1, so does the population of type T_2 cells. Using the approximation of the D -posterior distribution at \hat{p}_n in both experiments, we present an estimation of the probability that $m > 1$, with respect to the D -posterior distribution at \hat{p}_n , in Table 1 of Subsection 4.1 of the Supplementary Material. Note that the aforementioned probability is greater in the second experiment and this could mean that the solution with the hormone is effective at promoting cell reproduction. Indeed, by comparing Tables 3 and 4, one also sees that the probability that a cell of type T_1 dies with no offspring is smaller in the second experiment whereas the probability of differentiating into a type T_2 cell is greater.

6.2 Simulated example

This example is a continuation of Example 1 introduced in Section 2 and its purpose is to show the behaviour of EDAP and MDAP estimators in presence of contamination, specifically gross error contamination model. Recall that the process starts with $Z_0 = 1$ individual, the offspring distribution is a geometric distribution with parameter $\theta_0 = 0.3$ and which is affected by outliers, as described in Example 1, and the control variables $\phi_n(k)$ follow Poisson distributions with mean λk , for each $k \in \mathbb{N}_0$ and each $n \in \mathbb{N}_0$, with $\lambda = 0.3$. It is important to mention that the geometric distribution is a natural offspring distribution to use in the context of branching processes, as explained in the Introduction. Observe that we consider an extreme contamination framework; indeed, in this process, each progenitor has exactly 11 offspring with probability 0.15 and, with probability 0.85, it has offspring according to the aforementioned geometric distribution. From a practical viewpoint, this might seem to be too much extreme example, however, its choice is motivated by the fact that it allows to illustrate appropriately the accuracy of the method in a contamination context.

In order to approximate the D -posterior density functions based on the Hellinger distance (HD) and the negative exponential disparity (NED), one can consider different approaches. One of them is to obtain a sample of such a density function by applying the Metropolis-Hastings algorithm, and use this sample to estimate the corresponding density function (this approach was used in Hooker and Vidyshankar (2014)). However, since an expression of the D -posterior density function is available (see (5)), we have opted for estimating both the D -posterior density function and the EDAP and MDAP estimators by using numerical methods. The reason for the choice of the latter approach is that it is computationally much faster than the former; this fact is especially remarkable when considering a disparity measure different from the KL disparity since in that case it is not possible to use a conjugate families of distributions (see Hooker and Vidyshankar (2014)). Specifically, to approximate the integrals involved in (5) and (7) we used the Newton-Cotes formulas implemented in the function `cotes()` of the library `pracma` (see Borchers (2019)). On the other hand, to compute the MDAP estimates we have made use of the function `nloptr()` of the package `nloptr` again. We also remark that the computational cost is not significantly increased when replacing the KL disparity with the HD or the NED. Indeed, it only takes 0.08, 0.22, and 0.11 seconds to estimate the D -posterior density and to compute the EDAP and MDAP estimates for the HD, the NED, and the KL, respectively, for a single generation. This clearly shows the advantage of using numerical methods over other algorithms such as Metropolis-Hastings or Gibbs sampler in this context. Moreover, the differences in the computational time are essentially caused by the number of operations involved in the definition of the HD, NED, and KL. Computations were performed on an Intel Core i7-8650U CPU running at 1.90 GHz with 32 GB of RAM.

First, in Figure 3 in Subsection 5.2 of the Supplementary Material we show the estimates of the HD-posterior density and NED-posterior density upon the sample z_{45}^* and a beta distribution with parameters $1/2$ and $1/2$ as a prior distribution. Recall that the choice of this prior distribution is motivated by the fact we are not assuming any prior knowledge about the offspring parameter. One can observe that both functions

are approximately centred around the true value of the offspring parameter and the EDAP and MDAP estimators based on both of them provide accurate estimates of this parameter: $\theta_{45}^{*HD} = 0.2960$ and $\theta_{45}^{+HD} = 0.2951$, in the case of HD, and $\theta_{45}^{*NED} = 0.2949$ and $\theta_{45}^{+NED} = 0.2936$, for the NED.

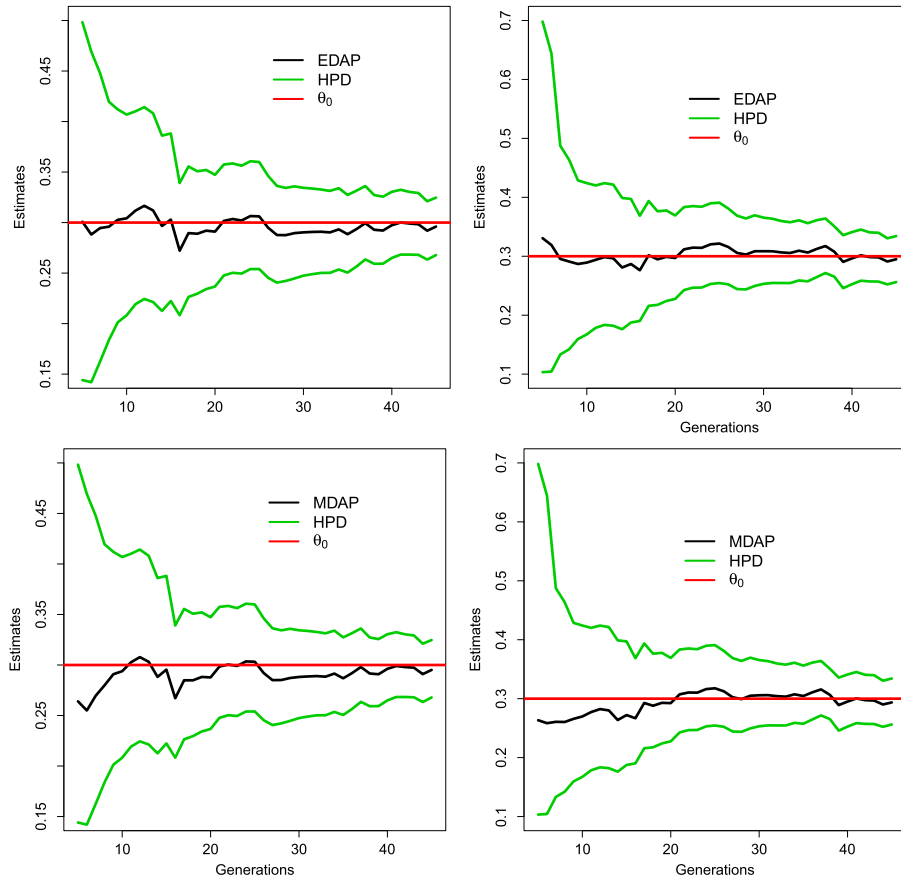


Figure 2: Evolution of EDAP and MDAP estimates (black line), and the HPD intervals (green lines) based on the HD (left) and the NED (right). Horizontal red lines represent the true value of θ_0 .

The strong consistency of EDAP and MDAP estimates based on HD and NED are illustrated in Figure 2. The evolution of HPD intervals is also plotted. These estimates are shown to be accurate in contrast to those plotted in Figure 1 (right). Moreover, by comparing the results obtained with the HD and the NED we conclude that there is no significant difference between them. This is not surprising due to the fact that each of them behaves as reasonably well as the other one against outlier contamination as our results in Section 5 state. The NED is, however, preferable under inlier contamination as explained in Lindsay (1994). The influence function plots and the robustness of HD

and NED are illustrated in Figure 3, for generations 25, 35 and 45. One can check a similar behaviour of the influence function plots for other values of n in Figure 6 in Subsection 5.3 of the Supplementary Material. We have also investigated the dependence of the influence function on the prior distribution as a part of the sensitivity study developed in Subsection 5.5 in the Supplementary Material. However, these graphs are omitted since no appreciable differences were observed.

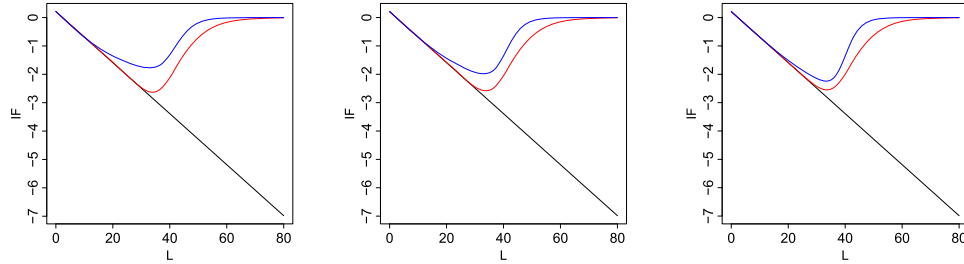


Figure 3: Plot of influence functions for EDAP estimator at p based on Kullback-Leibler (black line), Hellinger disparity (red line) and negative exponential disparity (blue line), $IF(L, \bar{T}_{25}, p)$ (left), $IF(L, \bar{T}_{35}, p)$ (middle), $IF(L, \bar{T}_{45}, p)$ (right), with $p = \mathbf{p}_{0.3}$ for the sample z_n^* , $n = 25, 35, 45$, with $\Delta_{24} = 95$, $\Delta_{34} = 156$, $\Delta_{44} = 308$, respectively.

Taking into account that the offspring mean and variance can be written as continuous functions of the offspring parameter, we have used the EDAP and MDAP estimates of the offspring parameter to obtain estimates of them. The evolution of $m(\theta_n^{*D})$, and $\sigma^2(\theta_n^{*D})$, for $n = 5, \dots, 45$, for $D \in \{HD, NED, KL\}$, are shown in Figure 4, whereas $m(\theta_n^{+D})$, and $\sigma^2(\theta_n^{+D})$, for $n = 5, \dots, 45$, for $D \in \{HD, NED, KL\}$, are shown in Figure 5, both of them in Subsection 5.2 of the Supplementary Material. Moreover, note that due to the continuity, these estimators prove to be strongly consistent estimators for the corresponding parameter. All the above empirical results, together with the theoretical results in previous sections, allow us to conclude that the HD and the NED are better choices for disparity measures than the likelihood disparity, because in absence of contamination, the three of them possess the same asymptotic properties, however under contamination, the HD and the NED are the only ones that gives robust estimates.

A sensitivity analysis was performed to determine the influence of the choice of the prior distribution in the D -posterior density and the corresponding point estimators. Since the parameter to be estimated lies in the interval $[0, 1]$, we use as prior distribution beta distributions with parameters (ρ, β) . Now, to apply the method, we take the parameters (ρ, β) belonging to the grid given by the Cartesian product of the set $\{n + 0.1 * k : n = 0, 1, \dots, 4; k = 1, 2, \dots, 10\}$ with itself. In Table 5 some results are summarized, especially those corresponding to prior distributions which are highly concentrated at extreme values far away from the parameter of interest. Despite this fact, one can observe a limited influence of the choice of the prior distribution on the EDAP and MDAP estimates. In addition, although the results are not shown here, it is interesting to mention that the effect of the prior distribution on the estimates decrease as the generation numbers increase (see Subsection 5.5 of the Supplementary Material for further details).

We also investigated other contamination models, namely, the mixture model for gross error by considering a geometric distribution with parameter $\theta_0 = 0.3$ and Poisson control distributions with mean $\lambda k = 0.3k$, with values of α and L different from 0.15 and 11, respectively. We emphasize here that the parameter $\lambda = 0.3$ is fixed and the role of k , as explained in the introduction, is to adapt to the number of parents in a given generation. Since the information in the sample until generation n depends on the mean growth rate, τ_m , it is a function of α and L . Thus, to make meaningful comparisons, we choose α and L that yield *similar* values of τ_m . As was already discussed, while Figure 1 illustrates the undesirable properties of EAP and MAP, the D -posteriors are unaffected by the contamination at different L and α . A summary of the behavior of EDAP estimates using KL, HD, and NED can be found in Figure 7 in Subsection 5.4 in the Supplementary Material. These results confirm that HD and NED are always preferable over KL. A similar observation also holds for MDAP.

We complete the illustration of the behaviour of the proposed methodology by considering the same mixture model for gross error of the previous offspring distribution and Poisson control distributions with mean $\lambda k = 0.45k$. In such a situation any value of $\alpha \in [0, 1]$ leads to a supercritical CBP (see Assumption 1). In Figure 8 in Subsection 5.4 in the Supplementary Material, similar behaviour of the estimates from the uncontaminated model ($\alpha = 0$) is illustrated; however, for the contaminated model ($\alpha = 0.01$) and as L increases, the EDAP and MDAP based on HD and NED are unaffected while that based on KL is considerably affected.

ρ	β	Prior mean	Prior variance	θ_{45}^{*HD}	θ_{45}^{*NED}	θ_{45}^{+HD}	θ_{45}^{+NED}
0.1	5	0.020	0.003	0.294	0.292	0.294	0.291
0.1	1	0.091	0.039	0.296	0.294	0.295	0.293
	2	0.286	0.026	0.296	0.294	0.295	0.293
	1	0.333	0.056	0.296	0.295	0.295	0.294
0.1	0.1	0.500	0.208	0.296	0.295	0.295	0.293
2.5	2.5	0.500	0.042	0.297	0.296	0.296	0.295
	2	0.667	0.056	0.297	0.297	0.296	0.295
	5	0.714	0.026	0.299	0.300	0.298	0.299
	1	0.909	0.039	0.296	0.296	0.296	0.295
	5	0.980	0.003	0.299	0.301	0.298	0.300

Table 5: EDAP and MDAP estimates for the HD and NED considering different beta distributions as prior distributions.

7 Concluding remarks

For controlled branching processes, this paper addresses the robust estimation of the parameters of the offspring distribution in a Bayesian context making use of disparity-based methods. We have assumed that the offspring distribution belongs to a general parametric family and the inference on the parameter of the reproduction law is based on the sample given by the entire family tree.

First, we have defined the D -posterior density, a density function which is obtained by replacing the log-likelihood in Bayes rule with a conveniently scaled disparity measure. The expectation and mode of the D -posterior density, referred to as EDAP and MDAP estimators, respectively, are proposed as Bayes estimators for the offspring parameter by emulating the point estimators under the squared error loss function or under $0 - 1$ loss function, respectively, for the posterior density. As a initial step for the analysis of the asymptotic and robustness properties of these estimators, their existence and measurability are studied. Moreover, sufficient conditions for the strong consistency and asymptotic normality of the EDAP and MDAP estimators, once they have been suitably centered and normalized, are provided.

In this paper we paid special attention to robust Bayesian inference for dependent tree-structured data. The EDAP and MDAP estimators, in the current context, are harder to analyze than the corresponding ones in the i.i.d. setting, due to the random sample size and the correlations induced by the random control function.

The analyses involve continuity and differentiability properties of Bayesian disparity functionals. These properties play an important role in the further development of robustness ideas in the context of the current data. We illustrate that the proposed estimators exhibit the robustness features when the number of generations increase.

Although the results are provided for a general class of disparity measures, we have focused our attention on the Hellinger distance and the negative exponential disparity in the examples presented. Contrary to the likelihood disparity, these disparity measures are known to provide robust estimators in an i.i.d. setting, while keeping desirable properties of estimators such as consistency and efficiency. Our results in the previous sections show that these features also hold in the branching framework, and thus, these disparities are preferable to the likelihood disparity. In the first example, we applied the methodology to real data from oligodendrocyte cell populations and estimated the main parameters of the process suggested to model such populations. The results in this example illustrate that in the absence of contamination, the properties of the EDAP and MDAP estimators are similar to traditional estimators based on likelihood disparity. The second example considers simulated data with the purpose of showing and analysing the robustness properties of the EDAP and MDAP estimators established in this paper. Thus, taken together both examples illustrate that EDAP and MDAP estimators, are efficient in a contamination-free context and robust to model misspecification and presence of outliers.

We finally remark that since our ultimate goal is to provide estimators of the offspring parameter, in both examples we have made use of numerical methods to approximate the posteriors instead of using a Markov Chain Monte Carlo methodology (as used in Hooker and Vidyshankar (2014)), leading to a reduction in the computational time.

Supplementary Material

Supplementary Material of “Robust estimation in controlled branching processes: Bayesian estimators via disparities” (DOI: [10.1214/20-BA1239SUPP](https://doi.org/10.1214/20-BA1239SUPP); .pdf). Appendix, proofs,

additional figures, simulated data, summaries of the sensitivity analysis performed, and the computational codes developed for the examples are provided in the Supplementary Material.

References

- Angelov, A. G. and Slavtchova-Bojkova, M. (2012). “Bayesian estimation of the offspring mean in branching processes: Application to infectious disease data.” *Computers & Mathematics with Applications*, 64: 229–235. MR2944806. doi: <https://doi.org/10.1016/j.camwa.2012.01.049>. 1011
- Berger, J. O. and Bernardo, J. M. (1992). “Ordered group reference priors with application to the multinomial problem.” *Biometrika*, 79(1): 25–37. MR1158515. doi: <https://doi.org/10.1093/biomet/79.1.25>. 1014, 1028
- Bissiri, P., Holmes, C., and Walker, S. (2017). “A general framework for updating belief distributions.” *Journal of the Royal Statistical Society: Serie B (Statistical Methodology)*, 78: 1103–1130. MR3557191. doi: <https://doi.org/10.1111/rssb.12158>. 1010
- Borchers, H. W. (2019). *pracma: Practical Numerical Math Functions*. R package version 2.2.5. URL <https://CRAN.R-project.org/package=pracma>. 1030
- Brown, L. D. and Purves, R. (1973). “Measurable Selections of Extrema.” *The Annals of Statistics*, 1(5): 902–912. MR0432846. 1018
- Catoni, O. (2004). *Statistical learning theory and stochastic optimization: Ecole d’Eté Probabilités de Saint-Flour XXXI*. Springer-Verlag Berlin Heidelberg. MR2163920. doi: <https://doi.org/10.1007/b99352>. 1010
- Corral, A., García-Millán, R., and Font-Clos, F. (2016). “Exact Derivation of a Finite-Size Scaling Law and Corrections to Scaling in the Geometric Galton-Watson Process.” *PLoS ONE*, 11(9): 1–17. 1011
- 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. 1017
- del Puerto, I. M., González, M., Gutiérrez, C., Martínez, R., Minuesa, C., Molina, M., Mota, M., and Ramos, A. (eds.) (2016). *Branching Processes and Their Applications*, volume 219 of *Lecture Notes in Statistics*. Springer International Publishing. MR3470191. doi: <https://doi.org/10.1007/978-3-319-31641-3>. 1009
- 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. doi: https://doi.org/10.1007/978-3-662-12788-9_7. 1009
- Ghosh, A. and Basu, A. (2016). “Robust Bayes estimation using the density power divergence.” *Annals of the Institute of Statistical Mathematics*, 68(2): 413–437. MR3464228. doi: <https://doi.org/10.1007/s10463-014-0499-0>. 1010

- Ghosh, A. and Basu, A. (2017). “General Robust Bayes Pseudo-Posterior: Exponential Convergence results with Applications.” *arXiv preprint arXiv:1708.09692*. 1010
- González, M., del Puerto, I. M., Martínez, R., Molina, M., Mota, M., and Ramos, A. (eds.) (2010). *Workshop on Branching Processes and Their Applications*, volume 197 of *Lecture Notes in Statistics – Proceedings*. Springer-Verlag. MR2761879. doi: <https://doi.org/10.1007/978-3-642-11156-3>. 1009
- González, M., del Puerto, I. M., and Yanev, G. P. (2018). *Controlled Branching Processes*. ISTE Ltd and John Wiley and Sons, Inc. 1010, 1013
- González, M., Martínez, R., and Mota, M. (2005). “On the unlimited growth of a class of homogeneous multitype Markov chains.” *Bernoulli*, 11(3): 559–570. MR2147775. doi: <https://doi.org/10.3150/bj/1120591189>. 1026
- 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. doi: <https://doi.org/10.1016/j.csda.2015.01.015>. 1014, 1018
- González, M., Minuesa, C., and del Puerto, I. (2017). “Minimum disparity estimation in controlled branching process.” *Electronic Journal of Statistics*, 11(1): 295–325. MR3606772. doi: <https://doi.org/10.1214/17-EJS1232>. 1010, 1018, 1019, 1020, 1021, 1022, 1024
- González, M., Minuesa, C., del Puerto, I., and Vidyashankar, A. N. (2020). “Supplementary Material of “Robust estimation in controlled branching processes: Bayesian estimators via disparities”.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/20-BA1239SUPP>. 1011
- Guttorp, P. (1991). *Statistical inference for branching processes*. John Wiley and Sons, Inc. MR1254434. 1010
- Haccou, P., Jagers, P., and Vatutin, V. (2005). *Branching processes: variation, growth and extinction of populations*. Cambridge University Press. MR2429372. 1009
- Hampel, F. R. (1974). “The influence curve and its role in robust estimation.” *Journal of the American Statistical Association*, 69(346): 383–393. MR0362657. 1021
- Holgate, P. and Lakhani, K. H. (1967). “Effect of offspring distribution on population survival.” *Bulletin of Mathematical Biophysics*, 29(4): 831–839. 1011
- Hooker, G. and Vidyshankar, A. N. (2014). “Bayesian model robustness via disparities.” *Test*, 23: 556–584. MR3252095. doi: <https://doi.org/10.1007/s11749-014-0360-z>. 1010, 1011, 1012, 1015, 1021, 1030, 1034
- Hyrien, O., Ambeskovic, I., Mayer-Proschel, M., Noble, M., and Yakovlev, A. (2006). “Stochastic modeling of oligodendrocyte generation in cell culture: model validation with time-lapse data.” *Theoretical Biology and Medical Modelling*, 3(1): 21. 1025
- Jiang, W. and Tanner, M. A. (2008). “Gibbs posterior for variable selection in high-dimensional classification and data mining.” *The Annals of Statistics*, 36(5): 2207–2231. MR2458185. doi: <https://doi.org/10.1214/07-AOS547>. 1010

- Kimmel, M. and Axelrod, D. E. (2015). *Branching processes in biology*. Springer-Verlag. MR3310028. doi: <https://doi.org/10.1007/978-1-4939-1559-0>. 1009
- 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. doi: <https://doi.org/10.1214/aos/1176325512>. 1016, 1017, 1018, 1031
- Miller, J. W. and Dunson, D. B. (2018). “Robust Bayesian Inference via Coarsening.” *Journal of the American Statistical Association*, 0(0): 1–13. MR4011766. doi: <https://doi.org/10.1080/01621459.2018.1469995>. 1010
- Park, C. and Basu, A. (2004). “Minimum disparity estimation: asymptotic normality and breakdown point results.” *Bulletin of Informatics and Cybernetics*, 36: 19–33. MR2139489. 1023
- 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. doi: [https://doi.org/10.1016/S0167-7152\(00\)00112-7](https://doi.org/10.1016/S0167-7152(00)00112-7). 1010, 1021
- Tse, H., Weaver, W., and Di Carlo, D. (2012). “Increased Asymmetric and Multi-Daughter Cell Division in Mechanically Confined Microenvironments.” *PLoS ONE*, 7: 1–8. 1025
- Warnes, G. R., Bolker, B., and Lumley, T. (2018). *gtools: Various R Programming Tools*. R package version 3.8.1. URL <https://CRAN.R-project.org/package=gtools>. 1028
- Yakovlev, A. Y., Stoimenova, V. K., and Yanev, N. M. (2008). “Branching Processes as Models of Progenitor Cell Populations and Estimation of the Offspring Distributions.” *Journal of the American Statistical Association*, 103(484): 1357–1366. MR2504197. doi: <https://doi.org/10.1198/016214508000000913>. 1012, 1025
- Ypma, J., Borchers, H. W., and Eddelbuettel, D. (2018). *Package ‘nloptr’*. R package version 1.2.1. URL <https://cran.r-project.org/web/packages/nloptr/nloptr.pdf>. 1028

Acknowledgments

The authors would like to thank the Associate Editor and the Reviewers for providing invaluable comments and suggestions which have significantly improved this paper. This research was partially developed while Carmen Minuesa was a visiting PhD student at the Department of Statistics, George Mason University, and she is grateful for the hospitality and collaboration.