

Summary of the work entitled *Bayesian robust estimation in controlled branching processes*

Carmen Minuesa Abril

1 Context

This work belongs to the field of Markov processes and more specifically, to the context of branching processes. These processes are one of the most remarkable probability models for the description of the dynamic of populations and since the appearance of the basic model, they have been extensively developed from simple to complex models which allow to represent many practical situations in a variety of fields such as biology, epidemiology, genetics, medicine, nuclear physics, demography, actuarial mathematics, and algorithm and data structures.

2 State of Art

Within the class of the branching processes, an appealing generalization of the classical model is the controlled branching process (CBP), on which this work is focussed. The distinctive characteristic of the aforementioned model is that the number of individuals with reproductive capacity in each generation is determined by a random control function. The dynamic of this process can be described as follows: each individual reproduces independently of the others according to the same probability law, called offspring distribution, and when the number of individuals in each generation is known, a random control mechanism determines the number of progenitors that participate in the subsequent reproduction phase.

The flexibility of this model let include many well-known branching processes as particular cases in this class, for example, the Bienaymè–Galton–Watson process itself, the branching process with immigration, with random migration, with immigration at state zero or with bounded emigration, the branching process with adaptive control or with continuous state space.

Most recent research on this field has been focussed on the inferential theory motivated by the strong relation between the value of the parameters of this process and its behaviour. In this sense, in a frequentist framework, relevant papers are [2], [11], [5], [6], [12], [7] and [8]. From a Bayesian standpoint, [10], [3] and [4] deserve to be mentioned.

In the context of branching processes, the study of robust estimators has been scarcely dealt with and all the results correspond to the frequentist framework. For the Bienaymé–Galton–Watson process, this issue is addressed by using weighted least trimmed estimation in [14] or by applying minimum Hellinger distance estimation in [13]. The results of last paper have been extended to the controlled branching process by applying minimum disparity estimation methodology in [8]. As a result, this work presents pioneer results for the problem of developing a robust procedure against outliers for the main parameters of the model from a Bayesian outlook.

In a general Bayesian context, different approaches have been proposed to tackle the problem of obtaining robust estimators when the data are suspected of being contaminated with large outliers (see [1]). A standard strategy is to postulate a heavy tailed distribution for the data, however, it entails a loss of precision in a free-contamination context. The same side effect is produced when assuming a mixture model for the data, where one of the mixture components is the contamination; this practice also presents an additional difficulty since the knowledge of the number of mixtures and of the type of distributions involved are required. Another procedure is to use prior distributions which penalize large values of the parameter. As an alternative solution to the problem, a methodology based on disparity theory has been proposed in [9] for samples made up by independent and identically distributed random variables.

3 Main contributions

This work presents an extension of the results given in [9] to the context of branching processes. Assuming a parametric framework for the offspring distribution and considering the entire family tree as sample, the main goal is to estimate the main parameter of the offspring distribution, which is called offspring parameter.

Firstly, for an arbitrary disparity measure D , we define a *D-posterior distribution* for the offspring parameter. This is motivated by the fact that the Kullback-Leibler divergence between the nonparametric maximum likelihood estimator of the offspring distribution and the parametric family to which we assume the offspring distribution belongs appears in the expression of the log-likelihood function. This suggests replacing the Kullback-Leibler divergence in the Bayes rule with a disparity measure D that it is known to provide robust estimators in a frequentist framework. The resulting function is the so-called *D-posterior density function* and its expectation is proposed as a point estimator of the offspring parameter, which is called *expectation a D-posteriori (EDAP) estimator*. In relation to its properties, we establish conditions for:

- Its existence.
- Its strong consistency.

- Its limiting behaviour compared to the minimum disparity estimator of the offspring parameter (see [8] for its definition).
- The approximation of its limiting probability distribution by a normal distribution having been suitably normalized.

Besides the asymptotic features of the EDAP estimator, we also establish an approximation of the D -posterior density function by the density function of certain normal distribution.

For the analysis of the robustness properties of the proposed estimator, we introduce the *EDAP functions* and provide results concerning:

- Their existence.
- Their limiting behaviour compared to the disparity functional of the offspring parameter (see [8] for further details).

Regarding the robustness qualities, we analyse classical measures and a new notion introduced in [9], which we adapt to our context, the *asymptotic breakdown point*. In particular, we establish results for:

- The limiting behaviour of the α -influence curves.
- The boundedness of the influence curve.
- The breakdown point.
- The asymptotic breakdown point.

Finally, despite we provide the results for a large class of disparity measures, we focus our attention on the Hellinger distance and the negative exponential disparity to illustrate the accuracy of the proposed estimators by way of a simulated example using the statistical software and programming environment R.

Although we do not show the results in the present work given its extension, it is worthwhile to mention that, apart from the EDAP estimator, we have studied an estimator of the offspring parameter defined as the mode of the D -posterior density, obtaining similar results to those for the EDAP estimator.

References

- [1] J.O. Berger. An overview of robust Bayesian analysis. *Test*, 3:5–124, 1994.

- [2] J. P. Dion and B. Essebbar. On the statistics of controlled branching processes. *Lecture Notes in Statistics*, 99:14–21, 1995.
- [3] M. González, C. Gutiérrez, R. Martínez, and I. del Puerto. 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, 2013.
- [4] M. González, C. Gutiérrez, R. Martínez, C. Minuesa, and I. del Puerto. *Branching Processes and their Applications (in press)*, volume 219 of *Lecture Notes in Statistics-Proceedings*, chapter Bayesian inference for controlled branching processes. Springer-Verlag, 2016.
- [5] M. González, R. Martínez, and I. del Puerto. Nonparametric estimation of the offspring distribution and mean for a controlled branching process. *Test*, 13:465–479, 2004.
- [6] M. González, R. Martínez, and I. del Puerto. Estimation of the variance for a controlled branching process. *Test*, 14:199–213, 2005.
- [7] M. González, C. Minuesa, and I. del Puerto. Maximum likelihood estimation and Expectation-Maximization algorithm for controlled branching processes. *Computational Statistics and Data Analysis*, 93:209–227, 2016.
- [8] M. González, C. Minuesa, and I. del Puerto. Minimum disparity estimation controlled branching process. *Electronic Journal of Statistics (submitted)*, 2016. arXiv:1511.06400.
- [9] G. Hooker and A.N. Vidyshankar. Bayesian model robustness via disparities. *Test*, 43:556–584, 2014.
- [10] R. Martínez, M. Mota, and I. del Puerto. On asymptotic posterior normality for controlled branching processes. *Statistics*, 43(4):367–378, 2009.
- [11] K. Mohan. Estimation of parameters in controlled branching process. *Stochastic Modelling and Applications*, 3:1–15, 2000.
- [12] T. N. Sriram, A. Bhattacharya, M. González, R. Martínez, and I. del Puerto. Estimation of the offspring mean in a controlled branching process with a random control function. *Stochastic Processes and their Applications*, 117:928–946, 2007.
- [13] T. N. Sriram and A. N. Vidyashankar. Minimum Hellinger distance estimation for supercritical Galton–Watson processes. *Statistics and Probability Letters*, 50:331–342, 2000.
- [14] V. Stoimenova, D. Atanasov, and N. Yanev. Simulation and robust modifications of estimates in branching processes. *Pliska Studia Mathematica Bulgarica*, 16:259–271, 2004.