Subject Section

ESTIpop: A computational tool to simulate and estimate parameters for continuous-time Markov branching processes

James P. Roney†, Jeremy Ferlic‡, Franziska Michor§-∥, and Thomas O. McDonald§-∥*

†Harvard College, Cambridge, MA 02138, USA  ‡Department of Data Sciences and Center for Cancer Evolution, Dana-Farber Cancer Institute, and Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA 02215, USA. §Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138. ∥The Broad Institute of Harvard and MIT, Cambridge, MA 02139, USA. ○The Ludwig Center at Harvard, Boston, MA 02215, USA.

†These authors contributed equally

*To whom correspondence should be addressed.

Abstract

Summary: ESTIpop is an R package designed to simulate and estimate parameters for continuous-time Markov branching processes with constant or time-dependent rates, a common model for asexually reproducing cell populations. Analytical approaches to parameter estimation quickly become intractable in complex branching processes. In ESTIpop, parameter estimation is based on a likelihood function with respect to a time series of cell counts, approximated by the Central Limit Theorem for multitype branching processes. Additionally, simulation in ESTIpop via approximation can be performed many times faster than exact simulation methods with similar results.

Availability: ESTIpop is available as an R-package on Github (https://github.com/michorlab/estipop).

Contact: michor@jimmy.harvard.edu and mcdonald@jimmy.harvard.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Branching processes, a subclass of stochastic processes, have been used extensively to model the growth and composition of reproducing populations (Haccou, 2005; Axelrod & Kimmel, 2015). Characterized by rates that govern growth, death, and mutation of various cell types, branching processes model complex cellular systems and hierarchies and have been employed to investigate the evolutionary dynamics of cancer, where differences in the fitness of cells, mutations conferring resistance or other traits, and competition between clones and cell types affect the trajectory of a population of cells (Durrett, 2015; McDonald & Michor, 2017). ESTIpop is an R package designed to work with experimental cell count data to estimate the parameters of and simulate continuous-time Markov branching processes (CTMBPs). Estimation of parameters in branching processes has important implications in cancer evolution and understanding how heterogeneity arises via inferring tumor fitness and mutation rates (Williams et al, 2018, Bozic et al., 2013). We derive an asymptotic likelihood function with respect to a time series of counts from each cell type in the population. This asymptotic likelihood function also applies to systems in which rates of death, mutation, and reproduction may vary as functions of time. This approach allows for the robust inference of models describing time-dependent cellular kinetics, such as those which result from circadian modulation of the cell cycle and time-dependent drug concentrations. As branching process models increase in complexity, exact analytical approaches become intractable, even for simple processes such as the birth-death process (Tavaré 2018). The use of asymptotic likelihood functions provides a method to perform estimation in these scenarios when initial population sizes are large, as is common in cell viability assays and other in vitro studies. Because of the need for cell count data rather than a surrogate such as tumor sizes, the methods provided serve primarily for in vitro studies and designing experiments where environmental factors can be better controlled.
ESTIpop also simulates arbitrary CTMBP exactly or by drawing from the multivariate normal distribution (MVN). This approach is much faster than exact simulation, especially for time-dependent models since the methods using acceptance-rejection algorithms can be prohibitively slow.

2 Methods

A CTMBP describes a system of independent individuals of various types who reproduce and/or die at rates that may vary over time. The process is Markov because an individual’s remaining lifespan is independent of its current age. Due to independence of individuals, a CTMBP can be viewed as a sum of processes initiated by independently growing ancestors. For large ancestor counts, the number of descendants is approximately normally distributed according to the Central Limit Theorem (CLT) (Yakovlev & Yanev, 2009). ESTIpop estimates rates in a CTMBP via maximum likelihood. Previous work uses approximate normality in branching processes to estimate the birth and death rates in a one-type, time-homogenous, birth-death process (Liu & Crawford, 2018). Here we show that the CLT holds for ancestors of different types, and the population converges to a MVN. This finding is noteworthy, as there are cases where the ancestor population comprises various types, such as pre-existing drug-resistant clones present in a tumor of otherwise sensitive cells. We also derive the moments of the time-inhomogenous multitype Markov branching process, extending previous work in the one-type case (Cohn & Herring, 1981). Using these results, we compute an asymptotic likelihood function with respect to a time series of population data by numerically solving the moments of the asymptotic MVN distribution describing each data point, and then summing the resulting MVN likelihoods across the data.

Simulation uses the Stochastic Simulation Algorithm, and time-dependent rates are incorporated into the SSA by the use of adaptive thinning (Gillespie, 1977; Lewis & Shedler, 1979). Although the SSA returns exact simulation results, it becomes computationally expensive in large systems or simulation times. ESTIpop also allows simulation from the approximate normal distribution, increasing the speed of simulation and requiring only single draws from a distribution rather than one for each replication event as in the SSA.

3 Results

Vignettes included in the supplementary materials validate our estimation procedures by demonstrating accurate parameter recovery from simulated data. We show that one-type, two-type, four-type, and time-inhomogenous models can be estimated using quantities of data that are realistically acquired in large-scale in vitro experiments.

Additionally, we include an application of estimation from experimental data concerning the time-dependence of cellular responses to palbociclib treatment. Analyzing data from the Harvard Medical School LINCS database, we find that cell growth shows distinctly time-dependent dynamics, which cannot be adequately modeled by a time-homogenous branching process. We propose a second-order step-response model to explain these dynamics, and parameterize it using maximum likelihood estimation via ESTIpop. The resulting model is able to recapitulate the essential temporal dynamics of the data, demonstrating the importance of time-inhomogeneity for properly modeling cellular phenomena (Fig 1C). The prediction simulations (Fig 1D) closely replicate the true values (Fig 1B). Further results are expanded on in the supplement.

4 Conclusion

ESTIpop estimates and simulates CTMBPs with applications for analyzing biological processes primarily for use in in vitro experiments. Normal approximations provide a quick alternative for simulating systems involving extremely large sizes or lengthy simulation times.

Acknowledgements

The authors thank Michael Nicholson and Justin Dean for their review of the theory and the entire Michor lab for insightful discussions.

Funding

This work was supported by the Dana-Farber Cancer Institute Physical Science-Oncology Center; grant number NIH U54CA193461.

Conflict of Interest: none declared.

References


