Published in final edited form as:

J Stat Plan Inference. 2011 July 1; 141(7): 2209–2227. doi:10.1016/j.jspi.2011.01.016.

Quasi- and pseudo-maximum likelihood estimators for discretely observed continuous-time Markov branching processes

Rui Chen¹ and Ollivier Hyrien^{1,*}

¹Department of Biostatistics and Computational Biology, University of Rochester Medical Center

Abstract

This article deals with quasi- and pseudo-likelihood estimation in a class of continuous-time multi-type Markov branching processes observed at discrete points in time. “Conventional” and conditional estimation are discussed for both approaches. We compare their properties and identify situations where they lead to asymptotically equivalent estimators. Both approaches possess robustness properties, and coincide with maximum likelihood estimation in some cases. Quasi-likelihood functions involving only linear combinations of the data may be unable to estimate all model parameters. Remedial measures exist, including the resort either to non-linear functions of the data or to conditioning the moments on appropriate sigma-algebras. The method of pseudo-likelihood may also resolve this issue. We investigate the properties of these approaches in three examples: the pure birth process, the linear birth-and-death process, and a two-type process that generalizes the previous two examples. Simulations studies are conducted to evaluate performance in finite samples.

Keywords

Continuous-time branching process; Conditional inference; Non-identifiability; Estimating equation; Optimality; Birth process; Birth-and-death process

1. Introduction

A vast body of literature is concerned with problems of statistical inference for branching processes. Many publications have investigated the Bienaymé-Galton-Watson process (see Guttorp (1991), Yanev (2008) and references therein, for instance). Much fewer have considered continuous-time branching processes, especially when the process is partially observed.

Kendall (1949), Darwin (1956), and Keiding (1974, 1975) investigated maximum likelihood estimation for the pure birth and the linear birth-and-death processes. Oh, Severo and Slivka (1991) proposed approximate maximum likelihood estimators for a class of pure birth processes. Becker and Hasofer (1997) investigated estimation for the linear birth-and-death process under a sampling schema slightly different than the one considered herein. Nedelman, Downs and Pharr (1985) constructed maximum likelihood estimators for a multi-

© 2011 Elsevier B.V. All rights reserved.

*ollivier_hyrien@urmc.rochester.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

type Bellman-Harris branching process model of the proliferation of mast cells. Maximum likelihood estimation remains, however, generally unattainable for multi-type Bellman-Harris branching processes. Hoel and Crump (1974) considered moment-based estimation using asymptotic approximations to the moments of the process. Boucher et al (1999) used a least squares estimator for a multi-type Bellman-Harris branching process describing the generation of oligodendrocytes *in vitro*. More recently, Monte Carlo approaches have been proposed for the multi-type Bellman-Harris process and one of its extensions. Zorin et al (2000) constructed a simulated maximum likelihood estimator using a stochastic approximation algorithm. Hyrien (2007) and Hyrien et al (2005a, 2005b, 2010a, 2010c) considered a pseudo-likelihood approach solely based on the first two moments of the process. Quasi-likelihood estimation was considered by Hyrien et al (2010). Chen et al (2010) investigated composite likelihood estimators. Hyrien and Zand (2008) and Hyrien, Chen and Zand (2010) proposed a method combining a mixture model and composite likelihood for CFSE-labeling data.

This paper deals with the methods of quasi- and pseudo-likelihood for discretely observed continuous-time Markov branching processes. Maximum likelihood estimation will generally outperform these alternative approaches, except when they coincide. These two methods are still worthy of consideration because maximum likelihood estimation may sometime be difficult to carry out, either because an expression for the likelihood function cannot be obtained in closed-form (this is particularly true for age-dependent branching processes) or because evaluation of the likelihood function is numerically challenging. Although Markov processes do not always provide realistic models, investigation in this setting remains of interest for several reasons, including: they provide useful tools for preliminary investigation of complex phenomena; explicit results can sometimes be obtained, thereby allowing potential pitfalls to be more easily identified. The methods of quasi- and pseudo-likelihood have been extensively studied in various contexts. These methods may give strongly consistent and asymptotically normal estimators under appropriate, often mild, regularity conditions. We refer to Heyde (1997) for an overview of quasi-likelihood, and to Gourieroux, Monfort and Trognon (1984) for a presentation of pseudo likelihood estimation. Heyde and Lin (1992) proposed quasi-likelihood estimators for a Bienaymé-Galton-Watson process with immigration. Klimko and Nelson (1978) and Jacob (2010) considered conditional least squares for general classes of stochastic processes including discrete time branching processes. Hyrien (2007) investigated a pseudo-maximum likelihood estimator for the multi-type Bellman-Harris branching process (see also Hyrien et al (2005a, 2005b, Hyrien et al (2010)).

The objective of this paper is to outline the main properties of quasi- and pseudo-likelihood estimation. Both methods rest solely on moments of the process, which can be computed with relative ease. They allow “conventional” and conditional inference, and offer reasonable alternatives to maximum likelihood estimation. Although quasi- and pseudo-maximum likelihood estimators are generally different, we show that they may be asymptotically equivalent on the set of non-extinction. We discuss possible connections with maximum likelihood estimation. In the case of quasi-likelihood for the pure birth process, this is accomplished by using the fact that the distribution of the process belongs to the linear exponential family. Unlike quasi-likelihood, the method of pseudo-likelihood as originally proposed by Gourieroux, Monfort and Trognon (1984) has not been shown to enjoy any optimality properties. We construct optimal pseudo-maximum likelihood estimators using the theory of estimating equations (Godambe, 1985). We show that the choice of the conditioning sigma-algebra may be important in making model parameters estimable when using the method of conditional quasi-likelihood. Large sample properties (consistency and asymptotic normality) of the proposed estimators are outlined in the general case, and discussed in detail in a few examples. These properties are investigated

when an increasing number of independent realizations of the process are observed at bounded numbers of time points, and when a single realization is observed at an increasing number of discrete time points. The finite sample performance of the estimators are investigated and compared using simulations.

This paper is organized as follows. The process is presented in Section 2, along with expressions for its expected value and covariance function. Section 3 describes the methods of quasi- and pseudo-likelihood. Section 4 considers application to three specific examples: the pure birth process, the linear birth-and-death process, and a two-type process generalizing the previous two examples.

2. A multi-type continuous-time Markov branching process

2.1. The process

The process under consideration describes a population consisting of K cell types, $K \in \mathbb{N}$. Without loss of generality, it begins at time 0 with a single ancestor cell of age zero and type 1. Upon completion of its lifespan, every type- k cell produces $\zeta_{k,1}$ type-1 cells, ..., $\zeta_{k,K}$ type- K cells, where $\zeta_k = (\zeta_{k,1}, \dots, \zeta_{k,K})$ is a random vector with probability distribution function $p_k(\mathbf{x}) = \mathbb{P}(\zeta_k = \mathbf{x})$. Denote the support of ζ_k by $\mathcal{S}_k$. Every type- k cell lives a random lifespan τ_k with cumulative distribution function $F_k(t) = \mathbb{P}(\tau_k \leq t)$. The random variables τ_k and ζ_k are usually assumed mutually independent (see Athreya and Ney, 1972), so that their joint distribution can be written as $\mathbb{P}(\tau_k \leq t, \zeta_k = \mathbf{x}) = p_k(\mathbf{x})F_k(t)$. In the present paper, we consider a natural generalization of this process, which allows the distribution of the lifespan τ_k to

depend on the progeny vector ζ_k , and we write $F_k(t) = \sum_{\mathbf{x} \in \mathcal{S}_k} p_k(\mathbf{x})F_k(t, \mathbf{x})$, where $F_k(t, \mathbf{x}) = \mathbb{P}(\tau_k \leq t | \zeta_k = \mathbf{x})$ denotes the conditional cumulative distribution function of τ_k given the type- k cell generates a progeny vector $\zeta_k = \mathbf{x}$ when completing its lifespan. The lifespan τ_k , given $\zeta_k = \mathbf{x}$, is assumed to follow an exponential distribution with parameter $\lambda_{k,\mathbf{x}}$. Last, every cell of the population evolves independently of every other cell.

This process is a particular case of a multi-type continuous-time branching process considered by Hyrien et al (2005a), which was built on a two-type process earlier proposed by Jagers (1975). It is also akin to Sevast'yanov (1971)'s process in that both processes allow the progeny vector and the cell lifespan to be dependent. The former process was used to analyze the division of the so-called O-2A progenitor cells and their differentiation into terminally differentiated oligodendrocytes by Hyrien et al (2005a), where dissimilar distributions were used to model the distributions of the time to occurrence of these two events. An analysis of experimental data on this cell system suggested the need for such a feature (Hyrien et al, 2005a), and time-lapse cinematography data provided further evidence in support of the proposed model extension (Hyrien et al, 2006). Of note, the above process reduces to the "usual" multi-type Markov process (Athreya and Ney, 1972) if $\lambda_{k,\mathbf{x}} = \lambda_k$ for every $\mathbf{x} \in \mathcal{S}_k$.

We assume that $p_k(\mathbf{x})$ and $\lambda_{k,\mathbf{x}}$ are functions of a finite dimensional parameter $\theta_0 \in \Theta \subset \mathbb{R}^p$, Θ compact. Any element of Θ will be denoted by θ . Our objective is to construct estimators for θ on the basis of partial observations of the above process. The data structure will be presented in section 3.1.

2.2. Moments of the process

For every $k, l = 1, \dots, K$, let $Z_k^{(l)}(t)$ be an $\mathbb{N}$ -valued stochastic process denoting the number of type- k cells at any time $t \geq 0$ arising from a single type- l cell born at $t = 0$, and introduce the vector $Z^{(l)}(t) = \{Z_1^{(l)}(t), \dots, Z_K^{(l)}(t)\}'$. We investigate estimators defined from marginal and

conditional mean and variance-covariance of $Z^{(l)}(t)$, and we present below expressions for these moments. Their derivation proceeds from arguments similar to those employed for the multi-type Bellman-Harris branching process (See Sevast'yanov and Chistyakov (1971) for instance). Technical details are therefore omitted.

Let $\mathbf{s} = (s_1, \dots, s_K)$, $\mathbf{x} = (x_1, \dots, x_K)$, $\mathbf{0} = (0, \dots, 0)_{1 \times K}$, and $\mathbf{1} = (1, \dots, 1)_{1 \times K}$. For every $k = 1, \dots, K$, let $\mathbf{e}_k$ be a $1 \times K$ vector having all elements equal to zero, except for a one in the k -th place. Define the probability generating functions (p.g.f.)

$$H^{(l)}(\mathbf{s}; t) = H^{(l)}(\mathbf{s}; t, \theta) = E_{\theta} \left\{ \prod_{k=1}^K s_k^{Z_k^{(l)}(t)} \right\}.$$

Remark—This function is parameterized by θ , but we shall use $H^{(l)}(\mathbf{s}; t)$ and $H^{(l)}(\mathbf{s}; t, \theta)$ interchangeably depending upon the context. The same will be true for the expectation and variance-covariance functions to be defined below. Also, we shall omit unnecessary sub- and superscripts when $K = 1$.

The functions $H^{(l)}(\mathbf{s}; t)$ satisfy the set of functional equations

$$H^{(l)}(\mathbf{s}; t) = s_l \left\{ 1 - \sum_{\mathbf{x} \in S_l} p_l(\mathbf{x}) F_l(t, \mathbf{x}) \right\} + \sum_{\mathbf{x} \in S_l} p_l(\mathbf{x}) \int_0^t H^{(1)}(\mathbf{s}; t - \tau)^{x_1} \times \dots \times H^{(K)}(\mathbf{s}; t - \tau)^{x_K} dF_l(\tau, \mathbf{x}). \quad (1)$$

Let $m_k^{(l)}(t) = E\{Z_k^{(l)}(t)\}$, $m_k^{(l)}(t|u) = E\{Z_k^{(l)}(t)|Z^{(l)}(u)\}$ and $m_k^{(l)}\{t|Z^{(l)}(u) \neq \mathbf{0}\} = E\{Z_k^{(l)}(t)|Z^{(l)}(u) \neq \mathbf{0}\}$

denote marginal and conditional expectations, and write $\mathbf{m}^{(l)}(t) = \{m_1^{(l)}(t), \dots, m_K^{(l)}(t)\}'$,

$\mathbf{m}^{(l)}(t|u) = \{m_1^{(l)}(t|u), \dots, m_K^{(l)}(t|u)\}'$, and

$\mathbf{m}^{(l)}(t|Z^{(l)}(u) \neq \mathbf{0}) = \{m_1^{(l)}(t|Z^{(l)}(u) \neq \mathbf{0}), \dots, m_K^{(l)}(t|Z^{(l)}(u) \neq \mathbf{0})\}'$. Similarly, let

$v_{jk}^{(l)}(t, u) = \text{cov}\{Z_j^{(l)}(t), Z_k^{(l)}(u)\}$, $v_{jk}^{(l)}(t|u) = \text{cov}\{Z_j^{(l)}(t), Z_k^{(l)}(u)|Z^{(l)}(u)\}$ and

$v_{jk}^{(l)}(t|Z^{(l)}(u) \neq \mathbf{0}) = \text{cov}\{Z_j^{(l)}(t), Z_k^{(l)}(u)|Z^{(l)}(u) \neq \mathbf{0}\}$ denote the marginal and conditional

covariances, and introduce the matrix notation $V^{(l)}(t, u) = [v_{jk}^{(l)}(t, u)]_{j,k=1, \dots, K}$,

$V^{(l)}(t|u) = [v_{jk}^{(l)}(t|u)]_{j,k=1, \dots, K}$ and $V^{(l)}(t|Z^{(l)}(u) \neq \mathbf{0}) = [v_{jk}^{(l)}(t|Z^{(l)}(u) \neq \mathbf{0})]_{j,k=1, \dots, K}$. Since

$m_k^{(l)}(t) = \partial H^{(l)}(\mathbf{s}; t) / \partial s_k|_{\mathbf{s}=\mathbf{1}}$, the expectations satisfy the system of functional equations

$$m_k^{(l)}(t) = \left\{ 1 - \sum_{\mathbf{x} \in S_l} p_l(\mathbf{x}) F_l(t, \mathbf{x}) \right\} \chi_{\{l=k\}} + \sum_{\mathbf{x} \in S_l} p_l(\mathbf{x}) \sum_{\alpha=1}^K x_{\alpha} \int_0^t m_k^{(\alpha)}(t-\tau) dF_l(\tau, \mathbf{x}), \quad |k, l=1, \dots, K \quad (2)$$

where $\chi_{\{l=k\}}$ denotes the indicator function. The systems (2) contains renewal-type equations, and explicit solutions can be obtained by solving the system iteratively. These solutions are linear combinations of certain functions taking the form

$$C(u; k_1, \dots, k_m; \mathbf{x}^{(k_1)}, \dots, \mathbf{x}^{(k_m)}) = F_{k_1}(\cdot, \mathbf{x}^{(k_1)}) * \dots * F_{k_m}(u, \mathbf{x}^{(k_m)})$$

for some $(k_1, \dots, k_m) \in \mathbb{N}^m$, $\mathbf{x}^{(k_1)}, \dots, \mathbf{x}^{(k_m)} \in \{1, \dots, K\}^{\otimes K}$, where

$$F_{k_1}(\cdot, \mathbf{x}^{(k_1)}) * F_{k_2}(u, \mathbf{x}^{(k_2)}) = \int_0^u F_{k_1}(u-v, \mathbf{x}^{(k_1)}) dF_{k_2}(v, \mathbf{x}^{(k_2)})$$

stands for the convolution of $F_{k_1}(\cdot, \mathbf{x}^{(k_1)})$ and $F_{k_2}(\cdot, \mathbf{x}^{(k_2)})$. The function $C(u; k_1, \dots, k_m; \mathbf{x}^{(k_1)}, \dots, \mathbf{x}^{(k_m)})$ is the c.d.f. of a General Erlang distribution (Johnson and Kotz, 1970) or of a Phase-type distribution, and it can be computed explicitly. An example shall be treated in Section 4.3.

To compute the variance-covariance function of the process, compute first the matrix of second order moments $\mathbb{E}\{Z^{(l)}(t)Z^{(l)}(u)'\}$. Let $U_{jk}^{(l)}(u) = \mathbb{E}\{Z_j^{(l)}(u)Z_k^{(l)}(u)\}$, and write $U_k^{(l)}(u) = \{U_{k1}^{(l)}(u), \dots, U_{kK}^{(l)}(u)\}$. The following Lemma is established by using the Markov Property:

Lemma 1—For every $t \geq u \geq 0$ and $l = 1, \dots, K$, we have

1. $m^{(l)}(t|u) = \sum_{k=1}^K Z_k^{(l)}(u) m^{(k)}(t-u);$
2. $m^{(l)}(t|Z^{(l)}(u) \neq \mathbf{0}) = m^{(l)}(t|u) / \{1 - H^{(l)}(\mathbf{0}; u)\},$
3. $V^{(l)}(t, u) = \sum_{k=1}^K m^{(k)}(t-u) U_k^{(l)}(u) - m^{(l)}(t) m^{(l)}(u)'$,
4. $V^{(l)}(t|u) = \sum_{k=1}^K Z_k^{(l)}(u) V^{(k)}(t-u, t-u),$
5. $V^{(l)}(t|Z^{(l)}(u) \neq \mathbf{0}) = V^{(l)}(t|u) / \{1 - H^{(l)}(\mathbf{0}; u)\},$

where $H^{(l)}(\mathbf{0}; u) = \mathbb{P}\{Z^{(l)}(u) = \mathbf{0}\}$.

To derive an expression for $U_{jk}^{(l)}(u)$, we have

$$U_{jk}^{(l)}(u) = \begin{cases} \left. \frac{\partial^2 H^{(l)}(\mathbf{s}; u)}{\partial s_k \partial s_j} \right|_{\mathbf{s}=\mathbf{1}} + m_k^{(l)}(u) & \text{if } k=j \\ \left. \frac{\partial^2 H^{(l)}(\mathbf{s}; u)}{\partial s_k \partial s_j} \right|_{\mathbf{s}=\mathbf{1}} & \text{otherwise,} \end{cases}$$

where the partial second order derivatives are the solutions to the system of integral equations

$$\left. \frac{\partial^2 H^{(l)}(\mathbf{s}; u)}{\partial s_k \partial s_j} \right|_{\mathbf{s}=\mathbf{1}} = L_{kj}^{(l)}(u) + \sum_{\mathbf{x} \in \mathcal{S}_l} p_l(\mathbf{x}) \sum_{\alpha=1}^K x_\alpha \int_0^u \frac{\partial^2 H^{(\alpha)}(\mathbf{s}; u-\tau)}{\partial s_k \partial s_j} \Big|_{\mathbf{s}=\mathbf{1}} dF_l(\tau, \mathbf{x}) \quad (3)$$

for $l = 1, \dots, K$, and where

$$L_{kj}^{(l)}(u) = \sum_{x \in S_l} p_l(x) \sum_{\alpha=1}^K \sum_{\beta=1}^K x_\alpha x_\beta \int_0^u m_k^{(\alpha)}(u-\tau) m_j^{(\beta)}(u-\tau) dF_l(\tau, x) - \sum_{x \in S_l} p_l(x) \sum_{\alpha=1}^K x_\alpha \int_0^u m_k^{(\alpha)}(u-\tau) m_j^{(\alpha)}(u-\tau) dF_l(\tau, x).$$

These equations also can be solved explicitly (see remark above for renewal equations).

3. Estimation

3.1. Data structure

Let n denote the number of independent realizations of the process observed during the experiment. For every $i = 1, \dots, n$, let $0 = t_{i0} < t_{i1} < \dots < t_{in_i} < \infty$ be $n_i + 1$ non-negative ordered time points. For every $j = 0, \dots, n_i$, define the $K \times 1$ vectors $Y_{ij} = (Y_{ij1}, \dots, Y_{ijK})'$, where Y_{ijK} , $k = 1, \dots, K$, denotes the number of type- k cells for the i -th realization of the process at time t_{ij} . By convention, the first cell shall be of type 1 such that $Y_{i0} = e_1'$. Set $Y_i = (Y_{i1}', \dots, Y_{in_i}')'$. We assume that the vectors Y_i , $i = 1, \dots, n$, are independent realizations of the vector $\{Z^{(1)}(t_{i1})', \dots, Z^{(1)}(t_{in_i})'\}'$. The two most common experimental designs are (i) when multiple realizations of the process are each observed at a single or a small number of time points, possibly distinct for different realizations; and (ii) when a single realization of the process is observed repeatedly over time.

Another situation that is of practical interest is when the process begins with an arbitrary, possibly large, number of cells. We do not consider this situation here, but the method generalizes to this setting by adapting the expression for the moments of the process. We refer to Yanev (1975), Dion and Yanev (1994), and Yakovlev and Yanev (2009) for related results. The estimators obtained when n independent realizations of the process are observed separately are not always identical to those obtained when the process starts with n ancestors, even when the times of observations are identical for every realization.

Let the $Kn_i \times 1$ vector $\mu_i(\theta) = \{m^{(1)}(t_{i1}; \theta)', \dots, m^{(1)}(t_{in_i}; \theta)'\}'$ and the $Kn_i \times Kn_i$ matrix

$$\Omega_i(\theta) = \begin{pmatrix} V^{(1)}(t_{i1}, t_{i1}; \theta) & \dots & V^{(1)}(t_{i1}, t_{in_i}; \theta) \\ \vdots & \ddots & \vdots \\ V^{(1)}(t_{in_i}, t_{i1}; \theta) & \dots & V^{(1)}(t_{in_i}, t_{in_i}; \theta) \end{pmatrix} \quad (4)$$

denote the expectation and the variance-covariance matrix of the number of cells counted in any realization of the process at the time points $t_{i1}, \dots, t_{in_i}$. Let $\mathcal{F} = \{\mathcal{F}_{ij}, j = 1, \dots, n_i, i = 1, 2, \dots\}$ denote any collection of sigma-algebras, and let $\mu_{ij}^{\mathcal{F}}(\theta) = E_\theta\{Z^{(1)}(t_{ij}) | \mathcal{F}_{i,j-1}\}$ and $v_{ij}^{\mathcal{F}}(\theta) = \text{Var}_\theta\{Z^{(1)}(t_{ij}) | \mathcal{F}_{i,j-1}\}$ denote associated conditional expectation and variance-covariance matrix. The most obvious choice for $\mathcal{F}$ is the natural filtration denoted by $\mathcal{F}_2 = \{\mathcal{F}_{ij} : \mathcal{F}_{ij} = \sigma\{Y_{i0}, \dots, Y_{ij}\}\}$, but other choices are sometimes warranted; for instance, we shall use the collection $\mathcal{F}_3 = \{\mathcal{F}_{ij} : \mathcal{F}_{ij} = \sigma\{Y_{ij} \neq \mathbf{0}\}\}$ to remediate a non-identifiability issue for the linear birth-and-death process. In what follows, we consider estimating θ on the basis of $Y_1, \dots, Y_n$ using the moment-based methods of quasi-and pseudo-likelihood.

3.2. Quasi-likelihood estimation

A quasi-likelihood estimator can be computed by solving the quasi-score equation $Q_1(\theta) = 0$, provided such a solution exists, where

$$Q_1(\theta) = \sum_{i=1}^n \frac{\partial \mu_i(\theta)'}{\partial \theta} \Omega_i(\theta)^{-1} \{Y_i - \mu_i(\theta)\} \quad (5)$$

denotes the quasi-score function. It is an optimal estimating function within the class of estimating functions

$$C_1 = \left\{ \sum_{i=1}^n a_i(\theta) \{Y_i - \mu_i(\theta)\}, \text{ where } a_i(\theta) \text{ is non-random} \right\}$$

(Godambe, 1985; Heyde, 1997). We shall denote the corresponding estimator by $\hat{\theta}_1$, and refer to it as a “conventional” estimator because $Q_1(\theta)$ does not involve conditional moments.

As an alternative, one could define conditional quasi-likelihood estimators using the class of estimating functions

$$C_2 = \left\{ \sum_{i=1}^n \sum_{j=1}^{n_i} \alpha_{ij}(\theta) \{Y_{ij} - \mu_{ij}^{\mathcal{F}}(\theta)\}, \text{ where } \alpha_{ij}(\theta) \text{ is } \mathcal{F}_{i,j-1}\text{-measurable} \right\}.$$

Under certain regularity conditions, an optimal estimating function within C_2 will be given by

$$Q^{\mathcal{F}}(\theta) = \sum_{i=1}^n \sum_{j=1}^{n_i} \frac{\partial \mu_{ij}^{\mathcal{F}}(\theta)'}{\partial \theta} v_{ij}^{\mathcal{F}}(\theta)^{-1} \{Y_{ij} - \mu_{ij}^{\mathcal{F}}(\theta)\}$$

(Godambe, 1985; Heyde, 1997). In what follows, we denote the quasi-likelihood function and estimator corresponding to the natural sigma-algebra $\mathcal{F}_2$ by $Q_2(\theta)$ and $\hat{\theta}_2$, and the ones corresponding to $\mathcal{F}_3$ by $Q_3(\theta)$ and $\hat{\theta}_3$. It is not difficult to see that the quasi-score function $Q_2(\theta)$ reduces to

$$Q_2(\theta) = \sum_{i=1}^n \sum_{j=1}^{n_i} A_{ij}(\theta) B_{ij}(\theta)^{-1} C_{ij}(\theta),$$

where

$$A_{ij}(\theta) = \sum_{k=1}^K Y_{i,j-1,k} \frac{\partial m^{(k)}(t_{ij} - t_{i,j-1}, \theta)}{\partial \theta},$$

$$B_{ij}(\theta) = \sum_{k=1}^K Y_{i,j-1,k} V^{(k)}(t_{ij} - t_{i,j-1}, \theta),$$

and

$$C_{ij}(\theta) = Y_{ij} - \sum_{k=1}^K Y_{i,j-1,k} m^{(k)}(t_{ij} - t_{i,j-1}, \theta).$$

Moreover, when $K = 1$, the quasi-score function simplifies to

$$Q_2(\theta) = \sum_{i=1}^n \sum_{j=1}^{n_i} \frac{\partial m_1^{(1)}(t_{ij} - t_{i,j-1}; \theta)}{\partial \theta} \frac{\{Y_{ij} - Y_{i,j-1} m_1^{(1)}(t_{ij} - t_{i,j-1}; \theta)\}}{v_{11}^{(1)}(t_{ij} - t_{i,j-1}; \theta)}. \quad (6)$$

Since $\mathbb{E}_{\theta_0}\{Q_a(\theta_0)\} = 0$, $a = 1, 2, 3$, $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$ are expected to be at least weakly consistent under certain regularity conditions (Heyde, 1997). Furthermore, when $n \rightarrow \infty$, $\sup_{i=1,2,\dots} n_i < \infty$ and $\sup_{i,j=1,2,\dots} t_{ij} < \infty$, these estimators are also $\sqrt{n}$ -consistent and asymptotically Gaussian. It is well-known that the asymptotic variance-covariance matrix of $\hat{\theta}_1$, for instance, is given by

$$\sum_1(\theta_0) = \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{\partial \mu_i(\theta_0)}{\partial \theta} \Omega_i(\theta_0)^{-1} \frac{\partial \mu_i(\theta_0)}{\partial \theta} \right\}^{-1}$$

(Heyde, 1997). The asymptotic properties of $\hat{\theta}_1$ are more difficult to derive in the general case when $\sup_{i=1,2,\dots} n_i \rightarrow \infty$, and they depend on properties of the process. Those of the conditional quasi-likelihood estimators may be easier to derive than those of the conventional estimators by exploiting, for instance, the martingale difference property of $Q_2(\theta)$ and $Q_3(\theta)$. We refer to Dion and Keiding (1978) and to Feigin, (1976) for related discussions. Other techniques can be resorted to, including straightforward applications of existing asymptotic results for discrete-time branching processes using an embedded Bienaymé-Galton-Watson process, as proposed by Yanev (1975).

The classes $\mathcal{C}_1$ and $\mathcal{C}_2$ consist of estimating functions that are linear combinations of the Y_{ij} . In principle, classes of nonlinear functions of the data could also be used. For instance, when $K = 1$, θ could be estimated using the class of quadratic estimating functions

$$C_3 = \left\{ \sum_{i=1}^n \sum_{j=1}^{n_i} \alpha_{ij}(\theta) \left[\{Y_{ij} - \mu_{ij}^{\mathcal{F}}(\theta)\}^2 - v_{ij}^{\mathcal{F}}(\theta) \right], \text{ where } \alpha_{ij}(\theta) \text{ is } \mathcal{F}_{i,j-1}\text{-measurable} \right\}.$$

Alternative classes could be easily constructed either by using higher order moments of the data or by combining C_2 and C_3 , for instance. The optimal quadratic estimating equation within C_3 is the one for which the weights are given by

$$\alpha_{ij}(\theta) = \left[-2 \frac{\partial \mu_{ij}^{\mathcal{F}}(\theta)}{\partial \theta} \{Y_{ij} - \mu_{ij}^{\mathcal{F}}(\theta)\} - \frac{\partial v_{ij}^{\mathcal{F}}(\theta)}{\partial \theta} \right] \text{Var} \left[\{Y_{ij} - \mu_{ij}^{\mathcal{F}}(\theta)\}^2 | \mathcal{F}_{i,j-1} \right]^{-1}.$$

These weights require knowledge of the third and fourth order conditional moments of the process. This may render estimation less robust and not particularly precise, in addition to being more complex. The approach becomes justified when the classes C_1 and C_2 do not allow estimation of all parameters (see section 4.2 on the birth-and-death process for an example).

When $\mathcal{F} = \mathcal{F}_2$, the quasi-likelihood function involves only first- and second-order moments of the process. While robust with respect to misspecification of higher order moments, the method does not always permit estimation of all parameters. Choosing a different $\mathcal{F}$ may help with resolving identifiability issues, but possibly at the expense of robustness. For instance, the birth and the death rates of the linear birth-and-death process cannot be estimated using $\mathcal{F} = \mathcal{F}_2$, but they become estimable using $\mathcal{F} = \mathcal{F}_3$ (see section 4.2). As noted in Lemma 1, the conditional expectation of the process depends on the probability of extinction, which the model has to specify correctly for the estimating function to remain unbiased.

3.3. Pseudo- and conditional pseudo-maximum likelihood estimation

As an alternative, estimation can proceed from the method of pseudo-likelihood (Gourieroux, Monfort and Trognon, 1984; Concordet and Nunez, 2002; Hyrien et al, 2005a, 2005b, 2010; Hyrien, 2007). A (conventional) pseudo-maximum likelihood estimator can be defined as the vector θ_1 that maximizes the Gaussian pseudo-likelihood function

$$P_1(\theta) = - \sum_{i=1}^n [\{Y_i - \mu_i(\theta)\} \Omega_i(\theta)^{-1} \{Y_i - \mu_i(\theta)\} + \log |\Omega_i(\theta)|],$$

where $|\Omega_i(\theta)|$ denotes the determinant of the matrix $\Omega_i(\theta)$. When $n \rightarrow \infty$, $\sup_{i=1,2,\dots} n_i < \infty$ and $\sup_{i,j=1,2,\dots} t_{ij} < \infty$, we have that θ_1 converges in probability to θ_0 , and

$$\sqrt{n}(\tilde{\theta}_1 - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1}),$$

where $I(\theta_0) = \lim_{n \rightarrow \infty} \text{Var}_{\theta_0} \{n^{-1} \partial P_1(\theta_0) / \partial \theta\}$ and $J(\theta_0) = \lim_{n \rightarrow \infty} \mathbb{E}_{\theta_0} \{ \partial^2 P_1(\theta_0) / \partial \theta \partial \theta' \} / n$ (Hyrien, 2007). The asymptotic behavior of θ_1 is more difficult to study in the general case when $n_i \rightarrow \infty$, and it depends on properties of the process and of the experimental design.

A wider class of pseudo-maximum likelihood estimators can be constructed by using estimating functions that rely on the quadratic exponential family. Gourieroux, Monfort and Trognon (1984) proposed the pseudo-likelihood function $P(\theta) = \sum_{i=1}^n H_i(\theta)$, where

$$H_i(\theta) = A_i(\theta) + B_i(Y_i) + C_i(\theta)Y_i + Y_i' D_i(\theta)Y_i,$$

for properly chosen functions $A_i(\theta)$, $B_i(Y_i)$, $C_i(\theta)$ and $D_i(\theta)$ that involve certain characteristics of the data (such as their first and second order moments). For instance, in the case of the Gaussian pseudo-likelihood, we have $A_i(\theta) = -\mu_i(\theta)' \Omega_i(\theta)^{-1} \mu_i(\theta) - \log |\Omega_i(\theta)|$, $B_i(Y_i) = 0$, $C_i(\theta) = 2\Omega_i(\theta)^{-1} \mu_i(\theta)$, and $D_i(\theta) = -\Omega_i(\theta)^{-1}$. Other pseudo-likelihood functions exist (including Poisson, gamma, multinomial), the choice of which can sometimes be guided by proximity to the actual distribution of the data. For instance, in the case of a pure birth process, it is well-known that the process, when normalized by its expectation, converges to an exponential distribution with parameter 1. Since, for an exponential distribution, when $n_i = 1$ and $K = 1$, we have $A_i(\theta) = -\log \mu_i(\theta)$, $B_i(Y_i) = 0$, $C_i(\theta) = 1/\mu_i(\theta)$, and $D_i(\theta) = 0$, θ can be estimated by maximizing the (exponential) pseudo-likelihood function

$$P_{\text{exp}}(\theta) = - \sum_{i=1}^n \left\{ \frac{Y_i}{\mu_i(\theta)} + \log \mu_i(\theta) \right\}, \quad (7)$$

or solving the associated estimating equation $\partial P_{\text{exp}}(\theta)/\partial \theta = 0$. It is not difficult to show that $\mathbb{E}_{\theta_0} \{ \partial P_{\text{exp}}(\theta_0)/\partial \theta \} = 0$, so that the corresponding estimator is expected to be consistent under certain regularity conditions (Heyde, 1997). One necessary condition is that θ can be identified based solely on the expectation of the process. Therefore, the choice of a pseudo-likelihood function should also be driven by identifiability considerations.

The above class of pseudo-maximum likelihood estimators can be enriched by considering conditional pseudo-maximum likelihood estimators. For instance, a conditional Gaussian pseudo-maximum likelihood estimator maximizes the conditional (Gaussian) pseudo-likelihood function

$$P^{\mathcal{F}}(\theta) = - \sum_{i=1}^n \sum_{j=1}^{n_i} \left[\{Y_{ij} - \mu_{ij}^{\mathcal{F}}(\theta)\}' v_{ij}^{\mathcal{F}}(\theta)^{-1} \{Y_{ij} - \mu_{ij}^{\mathcal{F}}(\theta)\} + \log |v_{ij}^{\mathcal{F}}(\theta)| \right].$$

The corresponding estimating equations are given by $\partial P^{\mathcal{F}}(\theta)/\partial \theta = 0$, where

$$\frac{\partial P^{\mathcal{F}}(\theta)}{\partial \theta} = - \sum_{i=1}^n \sum_{j=1}^{n_i} \left[-2 \frac{\partial \mu_{ij}^{\mathcal{F}}(\theta)}{\partial \theta} v_{ij}^{\mathcal{F}}(\theta)^{-1} \{Y_{ij} - \mu_{ij}^{\mathcal{F}}(\theta)\} - \{Y_{ij} - \mu_{ij}^{\mathcal{F}}(\theta)\}' v_{ij}^{\mathcal{F}}(\theta)^{-1} \frac{\partial v_{ij}^{\mathcal{F}}(\theta)}{\partial \theta} v_{ij}^{\mathcal{F}}(\theta)^{-1} \{Y_{ij} - \mu_{ij}^{\mathcal{F}}(\theta)\} + \text{tr} \left\{ v_{ij}^{\mathcal{F}}(\theta)^{-1} \frac{\partial v_{ij}^{\mathcal{F}}(\theta)}{\partial \theta} \right\} \right]. \quad (8)$$

It can be shown without much difficulty that $\partial P^{\mathcal{F}}(\theta)/\partial \theta$ is centered about zero, thereby yielding also consistent estimators. It also defines a sequence of martingale differences when $\mathcal{F}$ is a filtration. Let $P_2(\theta)$ and θ_2 denote the (Gaussian) pseudo-maximum likelihood function and estimator corresponding to $\mathcal{F} = \mathcal{F}_2$.

It is worth-noting that $P^{\mathcal{F}}(\theta)$ and $Q^{\mathcal{F}}(\theta)$ are related through

$$\frac{\partial P^{\mathcal{F}}(\theta)}{\partial \theta} = 2Q^{\mathcal{F}}(\theta) + \sum_{i=1}^n \sum_{j=1}^{n_i} \left[\{Y_{ij} - \mu_{ij}^{\mathcal{F}}(\theta)\}' v_{ij}^{\mathcal{F}}(\theta)^{-1} \frac{\partial v_{ij}^{\mathcal{F}}(\theta)}{\partial \theta} v_{ij}^{\mathcal{F}}(\theta)^{-1} \{Y_{ij} - \mu_{ij}^{\mathcal{F}}(\theta)\} - \text{tr} \left\{ v_{ij}^{\mathcal{F}}(\theta)^{-1} \frac{\partial v_{ij}^{\mathcal{F}}(\theta)}{\partial \theta} \right\} \right], \quad (9)$$

so that the corresponding pseudo-maximum and quasi-likelihood estimators will be asymptotically equivalent whenever the sum in the right-hand side of equation (9) is asymptotically negligible compared to $Q^{\mathcal{F}}(\theta)$. For instance, when $K = 1$ and $\mathcal{F}$ denotes the natural filtration, $\partial P_2(\theta)/\partial \theta$ simplifies to

$$\frac{\partial P_2(\theta)}{\partial \theta} = 2Q_2(\theta) + \sum_{i=1}^n \sum_{j=1}^{n_i} \left[\frac{\partial v_1^{(1)}(t_{ij} - t_{i,j-1}; \theta)/\partial \theta}{Y_{i,j-1} v_{11}^{(1)}(t_{ij} - t_{i,j-1}; \theta)^2} \{Y_{ij} - Y_{i,j-1} m_1^{(1)}(t_{ij} - t_{i,j-1}; \theta)\}^2 - \frac{\partial v_{11}^{(1)}(t_{ij} - t_{i,j-1}; \theta)/\partial \theta}{v_{11}^{(1)}(t_{ij} - t_{i,j-1}; \theta)} \right]. \quad (10)$$

Let $n_\infty = \#\{i : Y_{ij} \rightarrow \infty, \text{ as } t_{ij} \rightarrow \infty, i = 1, \dots, n\}$ denote the number of realizations of the process which increase indefinitely. Suppose that n_∞/n does not converge to zero, that

$\partial m_1^{(1)}(t_{ij} - t_{i,j-1}; \theta)/\partial \theta_r \neq 0, r = 1, \dots, \dim(\theta)$, and that $\exists 0 < c < C < \infty$ such that $\inf_{i,j} \{|$

$t_{ij} - t_{i,j-1}| \} > c$ and $\sup_{i,j} \{|t_{ij} - t_{i,j-1}|\} < C$. Since $Y_{i,j-1}^{-1/2} \{Y_{ij} - Y_{i,j-1} m_1^{(1)}(t_{ij} - t_{i,j-1}; \theta)\} = O_p(1)$, the above assumptions, together with equation (6) and additional, mild regularity conditions, imply that the sum in the right-hand side of equation (10) may be of smaller order than $Q_2(\theta)$, and that pseudo- and quasi-likelihood estimating functions will be asymptotically equivalent in such circumstances. In particular, the quasi- and pseudo-maximum likelihood estimators will be asymptotically equivalent whenever the method of quasi-likelihood can estimate consistently the entire vector θ (see section 4.1 on the pure birth process for an example). Partition now θ into θ_1 and θ_2 , and suppose that the quasi-score function $Q_2(\theta)$ can estimate consistently θ_1 but not θ_2 . One can show that the quasi- and pseudo-maximum likelihood estimators of θ_1 will be asymptotically equivalent, but the lack of estimability of θ_2 by quasi-likelihood does not necessarily imply that it cannot be estimated using the method of pseudo-likelihood. The birth-and-death process is an example where neither $Q_1(\theta)$ nor $Q_2(\theta)$ can estimate both the birth and the death rates, but the Gaussian pseudo-likelihood will (see Section 4.2). Thus the method of Gaussian pseudo-likelihood may be

preferable to quasi-likelihood when the latter suffers from non-identifiability issues. These properties should be established rigorously on a case by case basis.

The estimating equation associated with the pseudo-likelihood function $P(\theta)$ is a member of the class of estimating functions

$$C_{PL} = \left\{ \sum_{i=1}^n a_i(\theta) \frac{\partial H_i(\theta)}{\partial \theta}, \text{ where } a_i(\theta) \text{ is non-random} \right\},$$

where

$$\frac{\partial H_i(\theta)}{\partial \theta} = \frac{\partial A_i(\theta)}{\partial \theta} + \frac{\partial C_i(\theta)}{\partial \theta} Y_i + Y_i' \frac{\partial D_i(\theta)}{\partial \theta} Y_i.$$

The estimating function $\partial P(\theta)/\partial \theta$ does not usually have any optimality properties (unless it coincides with the score function). An optimal estimating equation within C_{PL} is given by

$$\sum_{i=1}^n \alpha_i^*(\theta) \partial H_i(\theta) / \partial \theta, \text{ where}$$

$$\alpha_i^*(\theta) = E_{\theta} \left\{ \frac{\partial^2 H_i(\theta)}{\partial \theta \partial \theta'} \right\} \text{Var}_{\theta} \left\{ \frac{\partial H_i(\theta)}{\partial \theta} \right\}^{-1}$$

(Godambe, 1985). It involves moments of order 3 and 4 whenever $\partial D_i(\theta)/\partial \theta \neq 0$. When applied to the pure birth process, the optimal exponential pseudo-likelihood approach will lead to maximum likelihood estimation. Optimal conditional pseudo-likelihood estimators can be constructed similarly.

4. Examples and simulations

4.1. The pure birth process

This first illustration considers the pure birth process (for which $K = 1$) where every cell divides into two new cells upon completion of its lifespan with probability one. Let $\lambda = 1/\theta$ denote the expected lifespan. It is well known that the marginal expectation and covariance function of the process $Z(t)$ are respectively given by $m(t; \theta) = e^{\theta t}$ and $v(t, u; \theta) = e^{\theta t} (e^{\theta u} - 1)$, for every $t \geq u \geq 0$ (Athreya and Ney, 1972). Therefore $\mu_i(\theta) = (e^{t_{i1}\theta}, \dots, e^{t_{in_i}\theta})'$, and

$$\Omega_i(\theta) = \begin{pmatrix} e^{t_{i1}\theta}(e^{t_{i1}\theta} - 1) & \dots & e^{t_{in_i}\theta}(e^{t_{i1}\theta} - 1) \\ \vdots & \ddots & \vdots \\ e^{t_{in_i}\theta}(e^{t_{i1}\theta} - 1) & \dots & e^{t_{in_i}\theta}(e^{t_{in_i}\theta} - 1) \end{pmatrix}.$$

The conditional expectation and covariance function of the process are given by $m(t|u; \theta) = Z(u)e^{\theta(t-u)}$ and $v(t|u; \theta) = Z(u)e^{\theta(t-u)}(e^{\theta(t-u)} - 1)$ respectively. Although the maximum likelihood estimator is easily computed here, it is of interest to investigate the behavior and performance of quasi- and pseudo-maximum likelihood estimators. The associated quasi-score functions are

$$Q_1(\theta) = \sum_{i=1}^n (t_{i1} e^{\theta t_{i1}}, \dots, t_{in_i} e^{\theta t_{in_i}}) \begin{pmatrix} e^{\theta t_{i1}} (e^{\theta t_{i1}} - 1) & \dots & e^{\theta t_{in_i}} (e^{\theta t_{i1}} - 1) \\ \vdots & \ddots & \vdots \\ e^{\theta t_{in_i}} (e^{\theta t_{i1}} - 1) & \dots & e^{\theta t_{in_i}} (e^{\theta t_{in_i}} - 1) \end{pmatrix}^{-1} \begin{pmatrix} Y_{i1} - e^{\theta t_{i1}} \\ \vdots \\ Y_{in_i} - e^{\theta t_{in_i}} \end{pmatrix}, \quad (11)$$

and

$$Q_2(\theta) = \sum_{i=1}^n \sum_{j=1}^{n_i} \frac{t_{ij} - t_{i,j-1}}{e^{\theta(t_{ij} - t_{i,j-1})} - 1} \{Y_{ij} - Y_{i,j-1} e^{\theta(t_{ij} - t_{i,j-1})}\}. \quad (12)$$

When the times of observation are equidistant, with $\Delta = t_{ij} - t_{i,j-1}$, it is not difficult to show that the conditional quasi-likelihood estimator solving the equation $Q_2(\theta) = 0$ is explicitly given by

$$\hat{\theta}_2 = \log \left(\frac{\sum_{i=1}^n \sum_{j=1}^{n_i} Y_{ij}}{\sum_{i=1}^n \sum_{j=1}^{n_i} Y_{i,j-1}} \right) / \Delta, \quad (13)$$

which is also the maximum likelihood estimator (Keiding, 1974). The above expression for $\hat{\theta}_2$ does not hold true when the sampling times are no longer equidistant, but the quasi-likelihood estimator $\hat{\theta}_2$ will still coincide with the maximum likelihood estimator. To prove this, notice that the log-likelihood function takes the form

$$L(\theta) = \sum_{i=1}^n \sum_{j=1}^{n_i} (Y_{ij} - Y_{i,j-1}) \log(1 - e^{-\theta(t_{ij} - t_{i,j-1})}) - Y_{i,j-1} \theta(t_{ij} - t_{i,j-1}) + f(Y_{ij}, Y_{i,j-1}), \quad (14)$$

where $f(\cdot, \cdot)$ denotes a function that does not depend on θ (Kendall, 1949; Keiding, 1974). The maximum likelihood estimator $\hat{\theta}_{mle}$ is defined as the solution to the score equation $\partial L(\theta) / \partial \theta = 0$, where

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{i=1}^n \sum_{j=1}^{n_i} (t_{ij} - t_{i,j-1}) \left\{ \frac{Y_{ij} - Y_{i,j-1}}{e^{\theta(t_{ij} - t_{i,j-1})} - 1} - Y_{i,j-1} \right\} = Q_2(\theta),$$

and it is therefore identical to $\hat{\theta}_2$ even when the sampling times are not equidistant. Another way of proving this result is by first noticing from equation (14) that the conditional distribution of $Z(t)$ belongs to the linear exponential family, and by concluding using Theorem 2 in Wedderburn (1974) that $\hat{\theta}_2 = \hat{\theta}_{mle}$. Proposition 1 states another property of $\hat{\theta}_1$ and $\hat{\theta}_2$ that holds true whether or not the sampling times are equidistant (See Appendix 1 for a proof.).

Proposition 1—For the pure birth process, we have $\hat{\theta}_1 = \hat{\theta}_2 = \hat{\theta}_{mle}$.

Write $\hat{\theta}$ in place of $\hat{\theta}_1$ and $\hat{\theta}_2$. Asymptotic properties for $\hat{\theta}$ when a single realization of the process is observed at an increasing number of equidistant time points ($n = 1$, $t_{1j} = j\Delta$, and n_1

$\rightarrow \infty$) follows directly from work by Dion (1974), Keiding (1974), Yanev (1975) and Heyde (1975). In particular, $\hat{\theta}$ is strongly consistent and asymptotically Gaussian when

normalized by $\sqrt{\sum_{j=1}^{n_1} Y_{1j}}$. Specifically, we have that

$$\sqrt{\sum_{j=1}^{n_1} Y_{1,j-1}} (\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{e^{\Delta\theta_0} - 1}{\Delta^2 e^{\Delta\theta_0}} \right) \quad (15)$$

(e.g., Keiding (1974), Theorem 5.2). When one observes an increasing number of realizations at uniformly bounded numbers of equidistant time points ($n \rightarrow \infty$, $\sup_{i=1,2,\dots} n_i = R < \infty$, and $t_{i,j} = j\Delta$), one can show that

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow \mathcal{N} \left(0, \frac{(e^{\Delta\theta_0} - 1)^2}{\Delta^2 e^{\Delta\theta_0} \sum_{r=1}^R \pi_r (e^{r\Delta\theta_0} - 1)} \right), \quad (16)$$

where π_r , $r = 1, \dots, R$, denotes the limiting proportion of realizations of the process observed at exactly $n_i = r$ time points (see Appendix 2 for a proof). When $R = 1$, the asymptotic variance of $\hat{\theta}$ in (16) reduces to that in (15), but the rates of convergence of the estimator remain different in the two cases.

The conventional pseudo-likelihood function $P_1(\theta)$ is constructed from the expressions given above for $\mu_i(\theta)$ and $\Omega_i(\theta)$. When each realization of the process is observed at a single time point, $P_1(\theta)$ simplifies to

$$P_1(\theta) = - \sum_{i=1}^n \left\{ \frac{(Y_{i1} - e^{\theta t_{i1}})^2}{e^{\theta t_{i1}} (e^{\theta t_{i1}} - 1)} + \log(e^{\theta t_{i1}} - 1) + \theta t_{i1} \right\}.$$

The conditional Gaussian pseudo-likelihood function using the collection of sigma-algebras $\mathcal{A}_2$ is given by

$$P_2(\theta) = - \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ \frac{(Y_{ij} - Y_{i,j-1} e^{\theta\Delta})^2}{Y_{i,j-1} e^{\theta\Delta} (e^{\theta\Delta} - 1)} + \log Y_{i,j-1} + \theta\Delta + \log(e^{\theta\Delta} - 1) \right\}.$$

The asymptotic properties of the pseudo-maximum likelihood estimators depend on the sampling schema. This is conveniently illustrated using $\hat{\theta}_2$. In the case where $n \rightarrow \infty$, $\sup_{i=1,2,\dots} n_i < \infty$ and Δ is fixed, we have that $\hat{\theta}_2$ converges to θ_0 in probability, and $\sqrt{n}(\hat{\theta}_2 - \theta_0)$ is asymptotically normal, centered about zero, and with variance that, although cumbersome, can be computed explicitly. The proof of the asymptotic normality relies on the independence of the Y_i , $i = 1, \dots, n$ and a Central Limit Theorem for i.i.d. r.v.s (e.g., see Van der Vaart, 2007). The case where a single realization of the process is observed at an increasing number of equidistant time points ($n = 1$, $t_{1j} = j\Delta$ and $n_1 \rightarrow \infty$) is treated in the following Proposition.

Proposition 2—Suppose $n = 1$ and $t_{1j} = \Delta j$, $j = 1, 2, \dots, n_1$. For the pure birth process, the conditional Gaussian pseudo-maximum and quasi-likelihood estimators $\hat{\theta}_2$ and $\tilde{\theta}_2$ are asymptotically equivalent when $n_1 \rightarrow \infty$. Specifically, we have that

$$\sqrt{\sum_{j=1}^{n_1} Y_{1,j-1}} (\tilde{\theta}_2 - \theta_0) = \sqrt{\sum_{j=1}^{n_1} Y_{1,j-1}} (\hat{\theta}_2 - \theta_0) + o_p(1).$$

Consequently,

$$\begin{aligned} \tilde{\theta}_2 &= \theta_0 + o_p(1), \\ \sqrt{\sum_{j=1}^{n_1} Y_{1,j-1}} (\tilde{\theta}_2 - \theta_0) &\xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{e^{\Delta\theta_0} - 1}{\Delta^2 e^{\Delta\theta_0}}\right), \end{aligned}$$

and $\tilde{\theta}_2$ is asymptotically fully efficient.

Proof—The pseudo-maximum likelihood estimator $\tilde{\theta}_2$ is defined as a solution to the equation $\partial P_2(\theta)/\partial \theta = 0$, where

$$\frac{\partial P_2(\theta)}{\partial \theta} = 2Q_2(\theta) + \sum_{j=1}^{n_1} \left\{ \frac{(Y_{1j} - Y_{1,j-1} e^{\theta\Delta})^2}{Y_{1,j-1}} g(\theta, \Delta) - h(\theta, \Delta) \right\}, \quad (17)$$

where $g(\theta, \Delta) = \Delta(2e^{\theta\Delta} - 1)/e^{\theta\Delta}(e^{\theta\Delta} - 1)^2$, and where $h(\theta, \Delta) = \Delta + \Delta e^{\theta\Delta}/(e^{\theta\Delta} - 1)$. Let $W \sim \exp(1)$ denote the almost sure limit of $Y_{1,n_1}/e^{n_1\Delta\theta_0}$ as $n_1 \rightarrow \infty$. It is well known that

$\sum_{j=1}^{n_1} Y_{1j} / \sum_{j=1}^{n_1} e^{j\Delta\theta_0} \xrightarrow{a.s.} W$ (Heyde, 1970; Dion, 1974). Therefore, we have

$$\begin{aligned} \frac{Q_2(\theta)}{\sum_{j=1}^{n_1} e^{j\Delta\theta_0}} &= \left(\frac{\sum_{j=1}^{n_1} Y_{1j}}{\sum_{j=1}^{n_1} e^{j\Delta\theta_0}} - \frac{\sum_{j=1}^{n_1} Y_{1,j-1}}{\sum_{j=1}^{n_1} e^{j\Delta\theta_0}} e^{\Delta\theta} \right) \frac{\Delta}{e^{\theta\Delta} - 1} \\ &= W(1 - e^{(\theta - \theta_0)\Delta}) \frac{\Delta}{e^{\theta\Delta} - 1} + o_p(1). \end{aligned}$$

Since $(Y_{1j} - Y_{1,j-1} e^{\theta\Delta}) / \sqrt{Y_{1,j-1}} = O_p(1)$ as $j \rightarrow \infty$, and

$\sum_{j=1}^{n_1} e^{j\Delta\theta_0} = e^{\Delta\theta_0}(e^{n_1\Delta\theta_0} - 1)/(e^{\Delta\theta_0} - 1) \sim e^{n_1\Delta\theta_0}$, we deduce, as $n_1 \rightarrow \infty$, that

$$\frac{1}{\sum_{j=1}^{n_1} e^{j\Delta\theta_0}} \sum_{j=1}^{n_1} \left\{ \frac{(Y_{1j} - Y_{1,j-1} e^{\theta\Delta})^2}{Y_{1,j-1}} g(\theta, \Delta) - h(\theta, \Delta) \right\} = o_p(1), \quad (18)$$

and that

$$\lim_{n_1 \rightarrow \infty} \frac{1}{\sum_{j=1}^{n_1} e^{j\Delta\theta_0}} \frac{\partial P_2(\theta)}{\partial \theta} = 2W(1 - e^{(\theta - \theta_0)\Delta}) \frac{\Delta}{e^{\theta\Delta} - 1}, \quad (19)$$

with the convergence holding uniformly in θ . Since θ_0 is the unique root of the right-hand side of equation (19), we deduce the convergence in probability of $\hat{\theta}_2$ to θ_0 .

To show that $\tilde{\theta}_2$ is asymptotically equivalent to $\hat{\theta}_2$, we will consider the asymptotic expansion

$$\sqrt{\sum_{j=1}^{n_1} Y_{1,j-1}} (\tilde{\theta}_2 - \theta_0) = \frac{1}{\sqrt{\sum_{j=1}^{n_1} Y_{1,j-1}}} \frac{\partial P_2(\theta_0)}{\partial \theta} / \left(\frac{1}{\sum_{j=1}^{n_1} Y_{1,j-1}} \left\{ \frac{\partial^2 P_2(\theta_0)}{\partial \theta^2} + o_p(1) \right\} \right).$$

We will prove that the above expansion is equivalent, up to an $o_p(1)$, to the asymptotic expansion for $\hat{\theta}_2$

$$\sqrt{\sum_{j=1}^{n_1} Y_{1,j-1}} (\hat{\theta}_2 - \theta_0) = \frac{Q_2(\theta_0)}{\sqrt{\sum_{j=1}^{n_1} Y_{1,j-1}}} / \left(\frac{1}{\sum_{j=1}^{n_1} Y_{1,j-1}} \left\{ \frac{\partial Q_2(\theta_0)}{\partial \theta} + o_p(1) \right\} \right).$$

Notice first that the second order derivative of $P_2(\theta)$ is given by

$$\frac{\partial^2 P_2(\theta_0)}{\partial \theta^2} = 2 \frac{\partial Q_2(\theta_0)}{\partial \theta} + \sum_{j=1}^{n_1} \left\{ \frac{(Y_{1j} - Y_{1,j-1} e^{\theta_0 \Delta})^2}{Y_{1,j-1}} \frac{\partial g(\theta_0, \Delta)}{\partial \theta} - 2(Y_{1j} - Y_{1,j-1} e^{\theta_0 \Delta}) \Delta e^{\theta_0 \Delta} g(\theta_0, \Delta) - \frac{\partial h(\theta_0, \Delta)}{\partial \theta} \right\}, \quad (20)$$

where

$$\frac{\partial Q_2(\theta_0)}{\partial \theta} = \sum_{j=1}^{n_1} \left\{ \frac{\Delta^2 e^{\theta_0 \Delta}}{(e^{\theta_0 \Delta} - 1)^2} (Y_{1j} - Y_{1,j-1} e^{\theta_0 \Delta}) - \frac{Y_{1,j-1} \Delta^2 e^{\theta_0 \Delta}}{e^{\theta_0 \Delta} - 1} \right\}.$$

It follows from (18) that

$$\frac{1}{\sum_{j=1}^{n_1} Y_{1,j-1}} \sum_{j=1}^{n_1} \frac{(Y_{1j} - Y_{1,j-1} e^{\theta_0 \Delta})^2}{Y_{1,j-1}} = e^{\Delta \theta_0} \frac{\sum_{j=1}^{n_1} e^{(j-1)\Delta \theta_0}}{\sum_{j=1}^{n_1} Y_{1,j-1}} \frac{1}{\sum_{j=1}^{n_1} e^{j\Delta \theta_0}} \sum_{j=1}^{n_1} \frac{(Y_{1j} - Y_{1,j-1} e^{\theta_0 \Delta})^2}{Y_{1,j-1}} = o_p(1).$$

We also have

$$\frac{1}{\sum_{j=1}^{n_1} Y_{1,j-1}} \sum_{j=1}^{n_1} (Y_{1j} - Y_{1,j-1} e^{\theta_0 \Delta}) = e^{\Delta \theta_0} \frac{\sum_{j=1}^{n_1} e^{(j-1)\Delta \theta_0}}{\sum_{j=1}^{n_1} Y_{1,j-1}} \frac{\sum_{j=1}^{n_1} (Y_{1j} - Y_{1,j-1} e^{\theta_0 \Delta})}{\sum_{j=1}^{n_1} e^{j\Delta \theta_0}} = o_p(1).$$

Therefore

$$\frac{1}{\sum_{j=1}^{n_1} Y_{1,j-1}} \frac{\partial^2 P_2(\theta_0)}{\partial \theta^2} = \frac{2}{\sum_{j=1}^{n_1} Y_{1,j-1}} \frac{\partial Q_2(\theta_0)}{\partial \theta} + o_p(1). \quad (21)$$

Using similar arguments applied to (17), we can also show that

$$\frac{1}{\sqrt{\sum_{j=1}^{n_1} Y_{1,j-1}}} \frac{\partial P_2(\theta_0)}{\partial \theta} = \frac{2Q_2(\theta_0)}{\sqrt{\sum_{j=1}^{n_1} Y_{1,j-1}}} + o_p(1). \quad (22)$$

We deduce from (17), (21) and (22) that

$$\sqrt{\sum_{j=1}^{n_1} Y_{1,j-1}} (\tilde{\theta}_2 - \theta_0) = \frac{Q_2(\theta_0)}{\sqrt{\sum_{j=1}^{n_1} Y_{1,j-1}}} / \left(\frac{1}{\sum_{j=1}^{n_1} Y_{1,j-1}} \left\{ \frac{\partial Q_2(\theta_0)}{\partial \theta} + o_p(1) \right\} \right) = \sqrt{\sum_{j=1}^{n_1} Y_{1,j-1}} (\hat{\theta}_2 - \theta_0) + o_p(1).$$

Proposition 2 follows from (15).

We deduce also from Propositions 1 and 2 that $\tilde{\theta}_2$ is asymptotically equivalent to the maximum likelihood estimator. The conditional pseudo-maximum and quasi-likelihood estimators will no longer be equivalent when $\sup_{i,j=1,2,\dots} t_{ij} < \infty$, even asymptotically when $n \rightarrow \infty$. We deduce from equation (7) that the (conventional) exponential pseudo-likelihood function, when independent realizations of the process are observed at a single time point each ($n_i = 1$), takes the form

$$P_{\exp}(\theta) = - \sum_{i=1}^n (Y_{i1} e^{-\theta t_{i1}} + \theta t_{i1}).$$

The associated estimating equation is given by

$$\frac{\partial P_{\exp}(\theta)}{\partial \theta} = \sum_{i=1}^n h_i(\theta)$$

where $h_i(\theta) = t_{i1}(Y_{i1} e^{-\theta t_{i1}} - 1)$. Since $E_{\theta}\{\partial h_i(\theta)/\partial \theta\} = -t_{i1}^2$ and $\text{Var}_{\theta}\{h_i(\theta)\} = t_{i1}^2 e^{-\theta t_{i1}} (e^{\theta t_{i1}} - 1)$, the optimal estimating equation constructed from an exponential pseudo-likelihood function is

$$\frac{\partial P_{\text{exp}}(\theta)}{\partial \theta} = \sum_{i=1}^n \frac{t_{i1} e^{\theta t_{i1}}}{e^{\theta t_{i1}} - 1} (Y_{i1} e^{-\theta t_{i1}} - 1),$$

which coincides with the quasi-likelihood function $Q_1(\theta)$ in this particular case.

Using simulations we compared the performance of the above estimators under a longitudinal design where the process was observed daily from day 1 to day 4. The true value of the mean time to division was $20 = 1/\theta_0$. As shown in Table 1, all estimators improved as the sample size increases. Quasi-likelihood estimation outperformed the Gaussian pseudo-likelihood estimation, as expected since quasi-likelihood is equivalent to maximum likelihood here, but moderately only.

4.2. The linear birth-and-death process

A variety of estimators can be constructed using the methods of quasi- and pseudo-likelihood applied to a linear birth-and-death process. Under this model, every cell, upon completion of its lifespan, either divides into two new cells with probability $p = \phi/(\phi + \rho)$ or dies (that is, disappears) with probability $1 - p = \rho/(\phi + \rho)$. The length of the lifespan is assumed to follow an exponential distribution with parameter $\phi + \rho$. The parameters ϕ and ρ are nonnegative and referred to as the birth and the death rates. Our objective is to estimate $\theta = (\phi, \rho)'$.

The p.g.f. of the process takes the form

$$H(s; t) = \begin{cases} \frac{\rho(s-1) - e^{t(\rho-\phi)}(\phi s - \rho)}{\phi(s-1) - e^{t(\rho-\phi)}(\phi s - \rho)} & \text{if } \phi \neq \rho \\ \frac{s - \phi s t + \phi t}{1 - \phi s t + \phi t} & \text{if } \phi = \rho \end{cases}$$

(Athreya and Ney, 1972) where the sub- and superscripts of Section 2.2 have been dropped to ease notation. The maximum likelihood estimator can be constructed using the probability distribution as given by Darwin (1956). Although the likelihood function takes an explicit form, we found that the calculation of the maximum likelihood estimator became numerically challenging as the size of the process got large, thereby warranting alternative approaches to be investigated.

Write $\zeta = \phi + \rho$ and $\delta = \phi - \rho$. Under this parameterization, the expectation of the process is given by $m(t; \zeta, \delta) = e^{t\delta}$, while its covariance, for every $t \geq u \geq 0$, is

$$v(t, u; \zeta, \delta) = \begin{cases} \frac{\zeta}{\delta} e^{t\delta} (e^{u\delta} - 1) & \text{if } \delta \neq 0 \\ u(\zeta + \delta) & \text{if } \delta = 0. \end{cases}$$

When $\delta \neq 0$, since $\partial m(t; \zeta, \delta) / \partial \zeta = 0$ and $\partial m(t; \zeta, \delta) / \partial \delta = t e^{t\delta}$, the quasi-score function $Q_1(\zeta, \delta)$ can be expressed as

$$Q_1(\zeta, \delta) = \frac{\delta}{\zeta} \sum_{i=1}^n \begin{pmatrix} 0 & \cdots & 0 \\ -t_{i1} e^{t_{i1}\delta} & \cdots & -t_{in_i} e^{t_{in_i}\delta} \end{pmatrix} \Gamma_i(\delta), \quad (23)$$

where

$$\Gamma_i(\delta) = \begin{pmatrix} e^{t_{i1}\delta}(e^{t_{i1}\delta} - 1) & \dots & e^{t_{i1}\delta}(e^{t_{in_i}\delta} - 1) \\ \vdots & \ddots & \vdots \\ e^{t_{in_i}\delta}(e^{t_{i1}\delta} - 1) & \dots & e^{t_{in_i}\delta}(e^{t_{in_i}\delta} - 1) \end{pmatrix}^{-1} \begin{pmatrix} Y_{i1} - e^{t_{i1}\delta} \\ \vdots \\ Y_{in_i} - e^{t_{in_i}\delta} \end{pmatrix}.$$

The solution to $Q_1(\zeta, \delta) = 0$, therefore, will not be unique, when it exists. Using $Q_2(\zeta, \delta)$ instead of $Q_1(\zeta, \delta)$ does not solve the problem, so that both estimating equations fail to estimate the birth and the death rates. We can however estimate their difference δ . Denoting by $\hat{\delta}_1$ and $\hat{\delta}_2$ the corresponding quasi-likelihood estimators, we have:

Proposition 3—For the birth-and-death process, we have that $\hat{\delta}_1 = \hat{\delta}_2$. Furthermore, when the sampling times are equidistant, we have that $\hat{\delta}_1 = \hat{\delta}_2 = \hat{\delta}_{mle}$, where

$$\hat{\delta}_{mle} = \log \left(\frac{\sum_{i=1}^n \sum_{j=1}^{n_i} Y_{ij}}{\sum_{i=1}^n \sum_{j=1}^{n_i} Y_{i,j-1}} \right) / \Delta, \quad (24)$$

where $\Delta = t_{ij} - t_{i,j-1}$.

Proof—The proof of Proposition 3 follows directly from results of Section 4.1 and by noticing that both quasi-score equations $Q_1(\zeta, \delta)$ and $Q_2(\zeta, \delta)$ can be equivalently expressed in terms of the quasi-score functions given in (11) and (12). The estimator $\hat{\delta}_{mle}$ in equation (24) is the maximum likelihood estimators for δ (Keiding, 1975).

To remediate the problem of non-estimability of ζ when using $Q_1(\zeta, \delta)$ and $Q_2(\zeta, \delta)$, we propose two alternative solutions. The idea behind the first solution is to construct a quasi-likelihood function where one conditions on sigma-algebras $\mathcal{F}_{ij}$ yielding conditional expectations $\mathbb{E}_\theta\{Z(t_{ij})|\mathcal{F}_{ij}\}$ such that the first order partial derivatives $\partial \mathbb{E}_\theta\{Z(t_{ij})|\mathcal{F}_{ij}\}/\partial \zeta$ and $\partial \mathbb{E}_\theta\{Z(t_{ij})|\mathcal{F}_{ij}\}/\partial \delta$ are not linearly dependent. One such example is provided when $\mathcal{F}_{ij}$ is the sigma-algebra generated by the event of non-extinction $\{Y_{ij} \neq 0\}$. The probability that the process becomes extinct before time t is given by

$$\mathbb{P}_\theta\{Z(t)=0\}=H(0;t)=\frac{-(\zeta-\delta)+(\zeta-\delta)e^{t\delta}}{-(\zeta+\delta)+(\zeta-\delta)e^{t\delta}}=1-\frac{\delta}{(\zeta-\delta)e^{t\delta}-(\zeta+\delta)}.$$

Writing $\Delta_{ij} = t_{ij} - t_{i,j-1}$, we deduce from Lemma 1 that

$$\begin{aligned} \mathbb{E}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1}) \neq 0\} &= \frac{(\zeta+\delta)e^{t_{ij}\delta} - (\zeta-\delta)e^{\Delta_{ij}\delta}}{2\delta} \\ \text{Var}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1}) \neq 0\} &= \frac{\zeta}{2\delta^2} \left\{ e^{t_{ij}\delta} - 1 \right\} \left\{ (\zeta+\delta)e^{t_{ij}\delta} - (\zeta-\delta)e^{\Delta_{ij}\delta} \right\}, \end{aligned}$$

and

$$\frac{\partial \mathbb{E}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1}) \neq 0\}}{\partial \zeta} = \frac{e^{t_{ij}\delta} - e^{\Delta_{ij}\delta}}{2\delta}.$$

If δ were known, an optimal estimating equation within C_2 for ζ would then take the form

$$Q_3(\zeta) = \sum_{i=1}^n \sum_{j=1}^{n_i} \frac{\partial \mathbb{E}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1}) \neq 0\}}{\partial \zeta} \frac{\{Y_{ij} - \mathbb{E}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1}) \neq 0\}\}}{\text{Var}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1}) \neq 0\}}.$$

The solution to the quasi-score equation $Q_3(\zeta) = 0$ does not appear to have any explicit expression, but can be approximated numerically. Since an optimal estimator for δ is readily available, we propose to replace δ by $\hat{\delta}$ when solving the above equation.

An alternative remedial measure to estimate both the birth and the death rates is to use an estimating equation that is a quadratic function of the data. Assuming again that δ is known, candidate estimating equations for estimating ζ would have the general form

$$Q_4(\zeta; \delta) = \sum_{i=1}^n \sum_{j=1}^{n_i} \alpha_{ij}(\zeta, \delta) \left[\{Y_{ij} - \mathbb{E}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1})\}\}^2 - \text{Var}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1})\} \right],$$

where $\alpha_{ij}(\zeta, \delta)$ are given weights. Set $\alpha_{ij}(\zeta, \delta) = 1$. When the sampling times are equidistant, the above equation yields the following estimator for ζ when δ is replaced by $\hat{\delta}$:

$$\hat{\zeta} = \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} (Y_{ij} - Y_{i,j-1} e^{\hat{\delta}\Delta})^2}{\left(\sum_{i=1}^n \sum_{j=1}^{n_i} Y_{i,j-1} \right) e^{\hat{\delta}\Delta} (e^{\hat{\delta}\Delta} - 1)} \hat{\delta}. \quad (25)$$

The optimal weight $\alpha_{ij}(\zeta, \delta)$ is given by

$$\alpha_{ij}(\zeta, \delta) = \frac{\partial \left[\{Y_{ij} - \mathbb{E}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1})\}\}^2 - \text{Var}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1})\} \right]}{\partial \zeta} \text{Var}_\theta\{(Z(t_{ij}) - \mathbb{E}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1})\})^2 | Z(t_{i,j-1})\}^{-1},$$

where

$$\frac{\partial \left[\{Y_{ij} - \mathbb{E}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1})\}\}^2 - \text{Var}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1})\} \right]}{\partial \zeta} = \frac{\partial \text{Var}_\theta\{Z(t_{ij})\mathbb{I}(t_{i,j-1})\}}{\partial \zeta} = \frac{Y_{i,j-1}}{\delta} e^{\delta \Delta_{ij}} (e^{\delta \Delta_{ij}} - 1).$$

The expression for the conditional variance $\text{Var}_\theta[(Z(t_{ij}) - \mathbb{E}_\theta\{Z(t_{ij})|Z(t_{i,j-1})\})^2 | Z(t_{i,j-1})]$ is given in Appendix 3. The associated quasi-likelihood estimator does not seem to have any explicit expression, but it is easily computed numerically. When the variance is replaced by 1, the above equation yields the following estimator for ζ when δ is replaced by $\hat{\delta}$:

$$\hat{\zeta} = \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} Y_{i,j-1} (Y_{ij} - Y_{i,j-1} e^{\hat{\delta}\Delta})^2}{\left(\sum_{i=1}^n \sum_{j=1}^{n_i} Y_{i,j-1}^2 \right) e^{\hat{\delta}\Delta} (e^{\hat{\delta}\Delta} - 1)} \hat{\delta}. \quad (26)$$

The conventional Gaussian pseudo-maximum likelihood estimator $(\tilde{\phi}_1, \tilde{\rho}_1)$ maximizes the pseudo-likelihood function $P_1(\theta)$. To show that this approach can estimate both ϕ and ρ , assume first that observations are collected at a single time point t . Notice that the pseudo-likelihood can be reparameterized as $\tilde{P}(m, \sigma^2) = -n \log \sigma^2 - \sum_{i=1}^n (Y_i - m)^2 / \sigma^2$, where $m = e^{t(\phi-\rho)}$ and $\sigma^2 = (\phi + \rho) e^{t(\phi-\rho)} \{e^{t(\phi-\rho)} - 1\} / (\phi - \rho)$. The function $P(m, \sigma^2)$ admits a unique maximum for $\tilde{P}(m, \sigma^2)$ attained at $\bar{m}_t = \sum_{i=1}^n Y_i / n$ and $\bar{\sigma}_t^2 = \sum_{i=1}^n (Y_i - \bar{m}_t)^2 / n$. A necessary and sufficient condition for ϕ and ρ to have a unique pseudo-maximum likelihood estimator is therefore that the system

$$\begin{cases} \bar{m}_t = e^{t(\phi-\rho)} \\ \bar{\sigma}_t^2 = \frac{\phi+\rho}{\phi-\rho} e^{t(\phi-\rho)} (e^{t(\phi-\rho)} - 1) \end{cases}$$

admits a single solution. Since $\bar{m}_t$ is always positive in the present context, it is easy to show that the pseudo-maximum likelihood estimators of ϕ and ρ always exist and are given by

$$\tilde{\phi}(t) = \frac{\log \bar{m}_t}{2t} \left\{ \frac{\bar{\sigma}_t^2}{\bar{m}_t(\bar{m}_t - 1)} + 1 \right\} \text{ and } \tilde{\rho}(t) = \frac{\log \bar{m}_t}{2t} \left\{ \frac{\bar{\sigma}_t^2}{\bar{m}_t(\bar{m}_t - 1)} - 1 \right\}. \quad (27)$$

This establishes estimability of ϕ and ρ by Gaussian pseudo-likelihood. Notice that, in equation (27), $\bar{\sigma}_t^2$ could be replaced by the unbiased estimator $n\bar{\sigma}_t^2/(n-1)$. Moreover, since $\bar{\sigma}_t^2 < \bar{m}_t(\bar{m}_t - 1)$ cannot be ruled out in finite samples, $\tilde{\rho}(t)$ may be negative, and a proper estimator is $\rho^+(t) = \rho(t) \vee 0$.

When the experimental design includes multiple time points, $t_1, \dots, t_m$, both ϕ and ρ can also be estimated by weighted averages of the $\tilde{\phi}(t_k)$ and $\tilde{\rho}(t_k)$: $\tilde{\phi}_\pi = \sum_{k=1}^m \pi_k \tilde{\phi}(t_k)$ and $\tilde{\rho}_\pi = \sum_{k=1}^m \pi_k \tilde{\rho}(t_k)$, where the π_k 's are fixed positive weights with $\sum_{k=1}^m \pi_k = 1$. In doing so, we avoid maximizing the pseudo-likelihood function using optimization algorithms. A simple choice for these weights is $\pi_k = 1/m$, $k = 1, \dots, m$, yielding

$$\tilde{\phi}_{ym} = \frac{1}{m} \sum_{k=1}^m \tilde{\phi}(t_k) \text{ and } \tilde{\rho}_{ym} = \frac{1}{m} \sum_{k=1}^m \tilde{\rho}(t_k).$$

A third estimator for ϕ and ρ is constructed using the maximum likelihood estimator of δ , as given in equation (24), which we plug into $P(\phi, \phi - \delta)$, and the birth rate is estimated by maximizing $P(\phi, \phi - \hat{\delta}_1)$ with respect to ϕ . The conditional Gaussian pseudo-likelihood $(\hat{\theta}_2$ and $\hat{\theta}_3)$ estimators can also be computed as presented previously.

We conducted a simulation study to evaluate the performance of the various estimators constructed above. The design used for these simulations was identical to the one considered for the pure birth process. Table 2 reports a summary of the results. The simulations suggest that the maximum likelihood estimator was the best of all. The second best estimator was the conditional Gaussian pseudo-likelihood estimator, which outperformed all quasi-likelihood estimators. While conditioning on the event of non-extinction did resolve the identifiability issue, the resulting estimator did not perform particularly well. Additional simulations (not shown) were carried out where the realizations of the process was observed over longer time periods. It is worth noting that we found maximum likelihood estimation numerically unstable when the size of the population grew too large, making numerical evaluation of the log-likelihood function challenging. No such problem occurred with the methods of pseudo- and quasi-likelihood, in which case they become particularly attractive.

4.3. A two-type process

We now consider a two-type process ($K = 2$) that extends the linear birth-and-death process. It begins with a single cell of type 1 and age zero. Upon completion of its lifespan, every type-1 cell either divides into two new cells of type 1 with probability p , or it produces a single type-2 cell, with probability $1 - p$. The progeny vector ζ_1 of type-1 cells is a 1×2 random vector with support $\mathcal{S}_1 = \{(2, 0), (0, 1)\}$. The time to division of any type-1 cell is described by an exponentially distributed random variable with parameter $\lambda_1 = \lambda_{1,(2,0)}$, whereas the time to differentiation of any type-1 cell follows an exponential distribution with parameter $\lambda_2 = \lambda_{1,(0,1)}$. To simplify the notations, we will denote their respective cumulative distribution functions by $G_1(t) = F_1(t, (2, 0))$ and $G_2(t) = F_1(t, (0, 1))$. In contrast, any type-2 cell has an infinite lifespan. Write $\theta = (p, \lambda_1, \lambda_2)'$ to denote the vector of unknown model parameters.

Let $Z_1(t)$ and $Z_2(t)$ denote the number of type-1 and the number of type-2 cells, respectively, at time $t \geq 0$, and set $Z(t) = \{Z_1(t), Z_2(t)\}'$. Let $m_1(t)$ and $m_2(t)$ denote the expectations of $Z_1(t)$ and $Z_2(t)$, and $v_{11}(t)$, $v_{22}(t)$ and $v_{12}(t)$ their respective variances and covariance.

It follows from equation (1) that the conditional p.g.f. $H^{(1)}(\mathbf{s}; t)$ satisfies

$$H^{(1)}(\mathbf{s}; t) = s_1 \{1 - pG_1(t) - (1-p)G_2(t)\} + p \int_0^t H^{(1)}(\mathbf{s}; t-\tau)^2 dG_1(\tau) + (1-p) \int_0^t H^{(2)}(\mathbf{s}; t-\tau) dG_2(\tau), \quad (28)$$

and $H^{(2)}(\mathbf{s}; t) = s_2$. Substituting this equation into (28) yields

$$H^{(1)}(\mathbf{s}; t) = s_1 \{1 - pG_1(t) - (1-p)G_2(t)\} + s_2(1-p)G_2(t) + p \int_0^t H^{(1)}(\mathbf{s}; t-\tau)^2 dG_1(\tau).$$

Differentiating with respect to $\mathbf{s}$ at $\mathbf{s} = \mathbf{1}$ yields

$$\frac{\partial H^{(1)}(\mathbf{1}; t)}{\partial s_1} = 1 - pG_1(t) - (1-p)G_2(t) + 2p \int_0^t \frac{\partial H^{(1)}(\mathbf{1}; t-\tau)}{\partial s_1} dG_1(\tau),$$

and

$$\frac{\partial H^{(1)}(\mathbf{1};t)}{\partial s_2} = (1-p)G_2(t)2p \int_0^t \frac{\partial H^{(1)}(\mathbf{1};t-\tau)}{\partial s_2} dG_1(\tau).$$

By solving the above renewal equations (Athreya and Ney, 1972), we deduce expressions for the expected numbers of type-1 and type-2 cells started from a type-1 cell:

$$m_1(t) = [1 - pG_1 - (1-p)G_2] * \sum_{l=0}^{\infty} (2p)^l G_1^{*l}(t) = \frac{1}{2} + \frac{1}{2} \sum_{l=0}^{\infty} (2p)^l G_1^{*l}(t) - m_2(t),$$

and

$$m_2(t) = (1-p) \sum_{l=0}^{\infty} (2p)^l G_1^{*l} * G_2(t).$$

Clearly, $G_1^{*l}(t)$ is the cumulative distribution function of an Erlang distribution with parameters l and λ_1 . When $\lambda_1 \neq \lambda_2$, an expression for $G_1^{*l} * G_2(t)$ can be computed explicitly

$$G_1^{*l} * G_2(t) = \frac{\lambda_2^l}{(\lambda_2 - \lambda_1)^l} G_2(t) - \sum_{i=0}^{l-1} \frac{\lambda_1 \lambda_2^i}{(\lambda_2 - \lambda_1)^{i+1}} G_1^{*(l-i)}(t),$$

and $m_1(t)$ and $m_2(t)$ can be computed explicitly.

To derive expressions for the variances and covariance of the process, we first compute the partial second order derivatives of $H^{(1)}(\mathbf{s}; t)$ with respect to $\mathbf{s}$ at $\mathbf{s} = \mathbf{1}$. It follows from equation (3) that for all $i, j = 1, 2$,

$$\frac{\partial^2 H^{(1)}(\mathbf{1};t)}{\partial s_i \partial s_j} = L_{ij}^{(1)}(t) + 2p \int_0^t \frac{\partial^2 H^{(1)}(\mathbf{1};t-\tau)}{\partial s_i \partial s_j} dG_1(\tau),$$

where

$$L_{ij}^{(1)}(t) = 2p \int_0^t m_i(t-\tau) m_j(t-\tau) dG_1(\tau).$$

We deduce that

$$\frac{\partial^2 H^{(1)}(\mathbf{1};t)}{\partial s_i \partial s_j} = \int_0^t m_i(t-\tau) m_j(t-\tau) \sum_{k=1}^{\infty} (2p)^k dG_1^{*k}(\tau), i, j=1, 2.$$

The expectations $m_1(t - \tau)$ and $m_2(t - \tau)$ in right-hand side of the above identity can be computed explicitly using expressions given before. The integrals, however, have quite tedious expressions, but they can be easily approximated using methods of numerical integration such as the Simpson or the Lobatto methods (Carnahan, Luther, Wilkes, 1969). The variances and covariance can then be deduced from the above expressions together with (2) in Lemma 1.

Define the first order partial derivatives $\partial m(t)/\partial \theta = \{\partial m_1(t)/\partial \theta, \partial m_2(t)/\partial \theta\}$, where

$$\frac{\partial m_k(t)}{\partial \theta} = \left\{ \frac{\partial m_k(t)}{\partial p}, \frac{\partial m_k(t)}{\partial \lambda_1}, \frac{\partial m_k(t)}{\partial \lambda_2} \right\}', k=1, 2$$

The entries of $\partial m(t, \theta)/\partial \theta$ are given by

$$\begin{aligned} \frac{\partial m_1(t)}{\partial p} &= -G_1(t) + \sum_{l=1}^{\infty} \left(2l(2p)^{l-1} G_1^{*l}(t) - (l+1)(2p)^l G_1^{*(l+1)}(t) \right) - \frac{\partial m_2(t)}{\partial p}, \\ \frac{\partial m_1(t)}{\partial \lambda_1} &= \sum_{l=0}^{\infty} (2p)^l \left(\frac{\partial G_1^{*l}(t)}{\partial \lambda_1} - p \frac{\partial G_1^{*(l+1)}(t)}{\partial \lambda_1} \right) - \frac{\partial m_2(t)}{\partial \lambda_1}, \\ \frac{\partial m_1(t)}{\partial \lambda_2} &= \frac{\partial m_2(t)}{\partial \lambda_2}, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial m_2(t)}{\partial p} &= -G_2(t) + \sum_{l=1}^{\infty} \left(2l(2p)^{l-1} - (l+1)(2p)^l \right) G_1^{*l} * G_2(t), \\ \frac{\partial m_2(t)}{\partial \lambda_1} &= (1-p) \sum_{l=0}^{\infty} (2p)^l \frac{\partial G_1^{*l} * G_2(t)}{\partial \lambda_1}, \\ \frac{\partial m_2(t)}{\partial \lambda_2} &= (1-p) \sum_{l=0}^{\infty} (2p)^l \frac{\partial G_1^{*l} * G_2(t)}{\partial \lambda_2}. \end{aligned}$$

The partial first order derivatives in the above quantities can be computed explicitly: (for $l \geq 1$)

$$\frac{\partial G_1^{*l} * G_2(t)}{\partial \lambda_1} = \frac{l\lambda_2^l}{(\lambda_2 - \lambda_1)^{l+1}} G_2(t) - \sum_{i=0}^{l-1} \left(\frac{\lambda_2^i (i\lambda_1 + \lambda_2)}{(\lambda_2 - \lambda_1)^{i+2}} G_1^{*(l-i)}(t) + \frac{\lambda_1 \lambda_2^i}{(\lambda_2 - \lambda_1)^{i+1}} \frac{\partial G_1^{*(l-i)}(t)}{\partial \lambda_1} \right),$$

and

$$\frac{\partial G_1^{*l} * G_2(t)}{\partial \lambda_2} = \frac{-l\lambda_1 \lambda_2^{l-1}}{(\lambda_2 - \lambda_1)^{l+1}} G_2(t) + \frac{\lambda_2^l}{(\lambda_2 - \lambda_1)^l} \frac{\partial G_2(t)}{\partial \lambda_2} + \sum_{i=0}^{l-1} \frac{\lambda_1 \lambda_2^{i-1} (i\lambda_1 + \lambda_2)}{(\lambda_2 - \lambda_1)^{i+2}} G_1^{*(l-i)}(t),$$

where

$$\frac{\partial G_i^{*l}(t)}{\partial \lambda_j} = -e^{\theta l i} \frac{t^l}{\lambda_i^{l+1}(l-1)!}$$

if $i = j$, and it is identical to zero otherwise.

We conducted a simulation study to compare the quasi- and pseudo-maximum likelihood estimators, for the two-type process. The study design was identical to that used in the previous simulation studies, except that every realization of the process was observed at a single time point ($n_i = 1$). Both quasi-likelihood approaches are therefore identical here. The results are shown in Table 3.

Unlike the simulation study on the pure birth process, this one indicates that both estimators perform similarly in the setting under consideration. The standard errors of either estimators for all three parameters decrease as the sample size increases. The estimator of the probability of division is essentially unbiased. The estimators of the mean time to division and of the mean time to differentiation are slightly biased, but these biases remain reasonable and decrease with increasing sample size. Both estimation methods are able to estimate the parameters of the lifespan distributions, even though these distributions are dissimilar.

5. Concluding remarks

We have investigated quasi- and pseudo-maximum likelihood estimation for continuous-time Markov branching processes. These methods offer generally applicable approaches to statistical inference, and they are particularly useful in situations where maximum likelihood estimation is difficult. Both approaches allow conventional and conditional inference, and can lead to optimal estimation by applying the theory of estimating equations (Godambe, 1895). When relying on first and second order moments of the process only, both methods will generally be robust with respect to misspecifications of higher order moments. Nonetheless, they sometimes lead to maximum likelihood estimation. Although the methods of quasi- and pseudo-likelihood are generally distinct, they may result in equivalent estimators under certain conditions (e.g., we have shown that they can be asymptotically equivalent on the set of non-extinction). When using linear combinations of the data only, quasi-likelihood estimators may be more prone to non-identifiability issues than Gaussian pseudo-likelihood estimators. Two mitigating measures were identified, including the use of (i) conditional quasi-likelihood functions where conditioning is done on appropriate sigma-algebras (such as the one generated by the event of non extinction), and (ii) quasi-likelihood functions involving non-linear functions of the data (e.g., quadratic estimating equations). We found that both options resulted in estimators with poor performance. In contrast, the conditional Gaussian pseudo-maximum likelihood estimator offered a better alternative that not only mitigated the non-identifiability issue but also demonstrated better properties in simulation studies. Direct derivation of maximum likelihood estimators for markov branching processes from likelihood function may be difficult. Keiding (1975) derived an expression for the maximum likelihood estimator for the difference of the birth and the death rates of a birth-and-death process indirectly using a modified sampling schema. In contrast, quasi-likelihood made it straightforward. Generally speaking, quasi- and maximum likelihood estimators are identical when the true distribution of the process belongs to the linear exponential family (Wedderburn, 1974). This result applies to the pure birth process, but not necessarily for more complex processes such as the birth-and-death process. Although not optimal, we found that pseudo-likelihood estimators can sometimes perform better than quasi-likelihood estimators.

Acknowledgments

This research was supported by grants R01 NS39511, R01 CA134839, R01 AI069351, and UL1 RR024160 from the National Institutes of Health (NIH).

Appendix 1

Proof of Proposition 1

Without loss of generality, assume $n = 1$, and write t_j for t_{1j} and Y_j for Y_{1j} . Define $Y = (Y_1, \dots, Y_{n_1})'$ and the diagonal matrix $Y^* = \text{diag}\{1, Y_1, \dots, Y_{n_1-1}\}$. Let $t = (t_1, \dots, t_{n_1})'$ and $\Delta t = (\Delta t_1, \dots, \Delta t_{n_1})'$, where $\Delta t_j = t_j - t_{j-1}$ for $j = 1, \dots, n_1$. Define the vector $M(t; \theta) = \{m_1(t_1; \theta), \dots, m_1(t_{n_1}; \theta)\}'$ and the $n_1 \times n_1$ diagonal matrix $D(\theta) = \text{diag}\{v(\Delta t_j, \Delta t_j; \theta)\}_{j=1, \dots, n_1}$.

The first quasi-score function defined in (5) takes the form

$$Q_1(\theta) = \frac{\partial M(t; \theta)}{\partial \theta} \Omega_1(\theta)^{-1} \{Y - M(t; \theta)\},$$

while the second one is given by

$$Q_2(\theta) = \frac{\partial M(\Delta t; \theta)}{\partial \theta} D(\theta)^{-1} \{Y - Y^* M(\Delta t; \theta)\}.$$

Define an $n_1 \times n_1$ matrix

$$\Psi(\theta) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ -e^{\Delta t_2 \theta} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & -e^{\Delta t_{n_1} \theta} & 1 \end{pmatrix}.$$

It is easy to derive that $Y - Y^* M(\Delta t; \theta) = \Psi(\theta) \{Y - M(t; \theta)\}$ by direct computation. Therefore, $Q_2(\theta)$ can be transformed into

$$Q_2(\theta) = \frac{\partial M(\Delta t; \theta)}{\partial \theta} D(\theta)^{-1} \Psi(\theta) \{Y - M(t; \theta)\}.$$

In order to prove that $Q_1(\theta) = Q_2(\theta)$, it suffices to show that

$$\frac{\partial M(t; \theta)}{\partial \theta} = \frac{\partial M(\Delta t; \theta)}{\partial \theta} D(\theta)^{-1} \Psi(\theta) \Omega_1(\theta). \quad (29)$$

Computing the i -th entry of the right-hand side of (29) yields

$$\begin{aligned}
\left\{ \frac{\partial M(\Delta; \theta)'}{\partial \theta} D(\theta)^{-1} \psi(\theta) \right\} \Omega(\theta)_i &= \begin{pmatrix} \frac{\Delta t_1}{e^{\Delta t_1 \theta} - 1} & -\frac{\Delta t_2 e^{\Delta t_2 \theta}}{e^{\Delta t_2 \theta} - 1} \\ \vdots & \vdots \\ \frac{\Delta t_{n-1}}{e^{\Delta t_{n-1} \theta} - 1} & -\frac{\Delta t_n e^{\Delta t_n \theta}}{e^{\Delta t_n \theta} - 1} \\ \frac{\Delta t_n}{e^{\Delta t_n \theta} - 1} & \end{pmatrix}' \begin{pmatrix} e^{t_1 \theta} (e^{t_1 \theta} - 1) \\ \vdots \\ e^{t_i \theta} (e^{t_i \theta} - 1) \\ \vdots \\ e^{t_n \theta} (e^{t_n \theta} - 1) \end{pmatrix} \\
&= (\Delta t_1 + \dots + \Delta t_i) e^{t_i \theta} - (e^{t_i \theta} - 1) \left(\frac{\Delta t_{i+1} e^{t_{i+1} \theta}}{e^{\Delta t_{i+1} \theta} - 1} - \frac{\Delta t_{i+1} e^{t_{i+1} \theta}}{e^{\Delta t_{i+1} \theta} - 1} + \dots + \frac{\Delta t_n e^{t_n \theta}}{e^{\Delta t_n \theta} - 1} - \frac{\Delta t_n e^{t_n \theta}}{e^{\Delta t_n \theta} - 1} \right) \\
&= t_i e^{t_i \theta},
\end{aligned}$$

which is the i -th entry of $\partial M(t; \theta)' / \partial \theta$. $\diamond$

Appendix 2

Proof of Asymptotic normality (16)

For the pure birth process observed at an increasing number of realizations at uniformly bounded numbers of equidistant time points ($n \rightarrow \infty$, $\sup_i n_i = R < \infty$, and $t_{ij} = j\Delta$), we have that

$$Q_2(\theta) = \sum_{i=1}^n \sum_{j=1}^{n_i} \frac{m(\Delta; \theta)}{v(\Delta, \Delta; \theta)} \{Y_{ij} - Y_{i,j-1} m(\Delta; \theta)\} = \sum_{r=1}^R \sum_{i=1}^{N_r} \sum_{j=1}^r \frac{m(\Delta; \theta)}{v(\Delta, \Delta; \theta)} \{Y_{ij} - Y_{i,j-1} m(\Delta; \theta)\},$$

where N_r denotes the number of realizations observed at r equidistant time points, such that $\sum_{r=1}^R N_r = n$. Since $\pi_r = \lim_{n \rightarrow \infty} N_r/n$, we deduce that

$$\frac{Q_2(\theta)}{\sqrt{n}} = \sum_{r=1}^R \frac{\sqrt{N_r}}{\sqrt{n}} \frac{1}{\sqrt{N_r}} \sum_{i=1}^{N_r} \sum_{j=1}^r \frac{m(\Delta; \theta)}{v(\Delta, \Delta; \theta)} \{Y_{ij} - Y_{i,j-1} m(\Delta; \theta)\}.$$

For $r = 1, \dots, R$, by invoking the independence of the $Y_i = (Y_{i1}, \dots, Y_{i,n_i})'$ and a Central Limit Theorem (Chow and Teicher, 2003) we obtain that

$$\frac{\sqrt{N_r}}{\sqrt{n}} \frac{1}{\sqrt{N_r}} \sum_{i=1}^{N_r} \left(\sum_{j=1}^r \frac{m(\Delta; \theta_0)}{v(\Delta, \Delta; \theta_0)} \{Y_{ij} - Y_{i,j-1} m(\Delta; \theta_0)\} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \pi_r V_r(\theta_0)), \text{ as } N_r \rightarrow \infty,$$

where

$$V_r(\theta_0) = \text{var}_{\theta_0} \left(\sum_{j=1}^r \frac{m(\Delta; \theta_0)}{v(\Delta, \Delta; \theta_0)} \{Y_{1j} - Y_{1,j-1} m(\Delta; \theta_0)\} \right).$$

Write $C_j(\theta_0) = Y_{1j} - Y_{1,j-1}m(\Delta; \theta_0)$. Clearly, $C_j(\theta_0)$ defines a martingale difference, so that $\mathbb{E}_{\theta_0}\{C_j(\theta_0)\} = 0$ and $\text{cov}_{\theta_0}\{C_{j_1}(\theta_0), C_{j_2}(\theta_0)\} = 0$ for $j_1 \neq j_2$. Therefore

$$\begin{aligned}
 V_r(\theta_0) &= \frac{m(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)^2} \sum_{j=1}^r \text{var}_{\theta_0} \left(\{Y_{1j} - Y_{1,j-1}m(\Delta; \theta_0)\} \right) \\
 &= \frac{m(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)^2} \sum_{j=1}^r \left[\mathbb{E}_{\theta_0} \{ \text{var}_{\theta_0} (Y_{1j} - Y_{1,j-1}m(\Delta; \theta_0)) \} + \text{var}_{\theta_0} \left(\mathbb{E}_{\theta_0} \{Y_{1j} - Y_{1,j-1}m(\Delta; \theta_0)\} \right) \right] \\
 &= \frac{m(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)^2} \sum_{j=1}^r \mathbb{E}_{\theta_0} \{ \text{var}_{\theta_0} (Y_{1j} - Y_{1,j-1}m(\Delta; \theta_0)) \} \\
 &= \frac{m(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)^2} \sum_{j=1}^r \mathbb{E}_{\theta_0} \{ Y_{1,j-1}^2 v(\Delta, \Delta; \theta_0) \} \\
 &= \frac{m(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)} \sum_{j=1}^r m((j-1)\Delta; \theta_0).
 \end{aligned}$$

Since $m(j\Delta; \theta) = e^{j\Delta\theta} = (e^{\Delta\theta})^j = m(\Delta; \theta)^j$, the expression for $V_r(\theta_0)$ becomes

$$V_r(\theta_0) = \frac{m(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)} \frac{1 - m(\Delta; \theta_0)^r}{1 - m(\Delta; \theta_0)}.$$

Therefore, we deduce that

$$\frac{\underline{Q}_2(\theta_0)}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{m(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)} \sum_{r=1}^R \pi_r \frac{1 - m(\Delta; \theta_0)^r}{1 - m(\Delta; \theta_0)} \right).$$

Consider next the first order derivative

$$\underline{Q}'_2(\theta) = \sum_{r=1}^R \sum_{i=1}^{N_r} \sum_{j=1}^r \left[\left(\frac{m'(\Delta; \theta)}{v(\Delta, \Delta; \theta)} \right)' \{Y_{ij} - Y_{i,j-1}m(\Delta; \theta)\} - Y_{i,j-1} \frac{m'(\Delta; \theta)^2}{v(\Delta, \Delta; \theta)} \right].$$

Notice that

$$\begin{aligned}
 \mathbb{E}_{\theta_0} \left\{ \left(\frac{m'(\Delta; \theta_0)}{v(\Delta, \Delta; \theta_0)} \right)' \{Y_{ij} - Y_{i,j-1}m(\Delta; \theta_0)\} - Y_{i,j-1} \frac{m'(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)} \right\} &= \mathbb{E}_{\theta_0} \left\{ -Y_{i,j-1} \frac{m'(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)} \right\} \\
 &= \frac{m'(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)} m(\Delta; \theta_0)^{j-1}.
 \end{aligned}$$

Using a Law of Large Numbers (Chow and Teicher, 2003), we deduce that

$$\begin{aligned}
\frac{Q'_2(\theta_0)}{n} &= \sum_{r=1}^R \frac{N_r}{n} \frac{1}{N_r} \sum_{i=1}^{N_r} \sum_{j=1}^r \left[\left(\frac{m'(\Delta; \theta_0)}{v(\Delta, \Delta; \theta_0)} \right)' \{Y_{ij} - Y_{i,j-1} m(\Delta; \theta_0)\} - Y_{i,j-1} \frac{m'(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)} \right] \\
&\xrightarrow{P} - \sum_{r=1}^R \pi_r \sum_{j=1}^r \frac{m'(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)} m(\Delta; \theta_0)^{j-1} \\
&= - \frac{m'(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)} \sum_{r=1}^R \pi_r \frac{1-m(\Delta; \theta_0)^r}{1-m(\Delta; \theta_0)}.
\end{aligned}$$

Therefore, we have $\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(\theta_0))$, where

$$\begin{aligned}
\Sigma(\theta_0) &= \left[- \frac{m'(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)} \sum_{r=1}^R \pi_r \frac{1-m(\Delta; \theta_0)^r}{1-m(\Delta; \theta_0)} \right]^{-2} \frac{m'(\Delta; \theta_0)^2}{v(\Delta, \Delta; \theta_0)} \sum_{r=1}^R \pi_r \frac{1-m(\Delta; \theta_0)^r}{1-m(\Delta; \theta_0)} \\
&= \frac{v(\Delta, \Delta; \theta_0)(1-m(\Delta; \theta_0))}{m'(\Delta; \theta_0)^2 \sum_{r=1}^R \pi_r (1-m(\Delta; \theta_0)^r)}.
\end{aligned}$$

Substituting $m(\Delta; \theta) = e^{\Delta\theta}$, $m'(\Delta; \theta) = \Delta e^{\Delta\theta}$, and $v(\Delta, \Delta; \theta) = e^{\Delta\theta}(e^{\Delta\theta} - 1)$ in the expression for $\Sigma(\theta_0)$, we finally obtain that

$$\sqrt{n}(\widehat{\theta} - \theta_0) \rightarrow \mathcal{N}\left(0, \frac{(e^{\Delta\theta_0} - 1)^2}{\Delta^2 e^{\Delta\theta_0} \sum_{r=1}^R \pi_r (e^{r\Delta\theta_0} - 1)}\right).$$

Appendix 3

Conditional second order moments of the birth-and-death process

Straightforward calculations show that

$$\begin{aligned}
&\text{Var}_{\theta}[(Z(t_{ij}) - \mathbb{E}\{Z(t_{ij}) | \mathcal{F}(t_{i,j-1})\})^2 | \mathcal{F}(t_{i,j-1})] \\
&= \text{Var}_{\theta}(Z(t_{ij})^2 | \mathcal{F}(t_{i,j-1})) \\
&\quad + 4\mathbb{E}_{\theta}(Z(t_{ij}) | \mathcal{F}(t_{i,j-1}))^2 \text{Var}_{\theta}(Z(t_{ij}) | \mathcal{F}(t_{i,j-1})) - 4\mathbb{E}_{\theta}(Z(t_{ij}) | \mathcal{F}(t_{i,j-1})) \text{Cov}_{\theta}(Z(t_{ij}), \\
&\quad Z(t_{ij})^2 | \mathcal{F}(t_{i,j-1})),
\end{aligned}$$

where

$$\begin{aligned}
\text{Var}_{\theta}(Z(t_{ij})^2 | \mathcal{F}(t_{i,j-1})) &= Z(t_{i,j-1})\{M_4(\Delta_{ij}) - M_2^2(\Delta_{ij})\} + 2Z(t_{i,j-1})(Z(t_{i,j-1}) - 1)\{M_2^2(\Delta_{ij}) - M_1^4(\Delta_{ij})\} + 4Z(t_{i,j-1})(Z(t_{i,j-1}) - 1)\{M_3(\Delta_{ij}) \\
&\quad - Z(t_{i,j-1})M_1(\Delta_{ij})\}, \\
\mathbb{E}_{\theta}(Z(t_{ij})^2 | \mathcal{F}(t_{i,j-1})) &= Z(t_{i,j-1})\{M_2(\Delta_{ij}) - M_1^2(\Delta_{ij})\}, \\
\text{Var}_{\theta}(Z(t_{ij}) | \mathcal{F}(t_{i,j-1})) &= Z(t_{i,j-1})\{M_2(\Delta_{ij}) - M_1^2(\Delta_{ij})\}, \\
\text{Cov}_{\theta}(Z(t_{ij})^2, Z(t_{ij}) | \mathcal{F}(t_{i,j-1})) &= Z(t_{i,j-1})M_3(\Delta_{ij}) + 3Z(t_{i,j-1})(Z(t_{i,j-1}) - 1)M_1(\Delta_{ij})M_2(\Delta_{ij}) + 6Z(t_{i,j-1})(Z(t_{i,j-1}) - 1)(Z(t_{i,j-1}) - 2)M_1^3(\Delta_{ij})
\end{aligned}$$

and where the $M_k(\Delta_{ij})$, $k = 1, 2, 3, 4$, are given in Appendix 4.

Appendix 4

Moments of the birth-and-death process

Let $M(s; t) = \mathbb{E}(e^{sZ(t)})$ denote the moment generating function of the process $Z(t)$. Put $A = A(s, t) = e^{t(\rho-\phi)}(\phi e^s - \rho) - \rho(e^s - 1)$ and $B = B(s, t) = e^{t(\rho-\phi)}(\phi e^s - \rho) - \phi(e^s - 1)$, and notice that $M(s; t) = A/B$ when $\rho \neq \phi$ (Allen, 2003). We also have the following useful identities: $A' = \partial A(s; t)/\partial s = e^{t(\rho-\phi)}(\phi s - \rho)e^s$, $B' = \partial B(s; t)/\partial s = e^{t(\rho-\phi)}(\phi s - \phi)e^s$, $A^{(k)} = \partial^k A(s; t)/\partial s^k = A^k$, and $B^{(k)} = \partial^k B(s; t)/\partial s^k = B^k$. Write $M^{(k)}(s, t) = \partial^k M(s; t)/\partial s^k$. A little algebra shows that

$$\begin{aligned} M^{(1)}(s, t) &= \frac{\frac{A}{B} - \frac{AB}{B^2}}{M^{(1)}(s, t) \left(1 - \frac{2B}{B}\right)} \\ M^{(2)}(s, t) &= \frac{M^{(1)}(s, t) \left(1 - \frac{2B}{B}\right)}{M^{(1)}(s, t) \left(1 - \frac{6B}{B} + \frac{6B^2}{B^2}\right)} \\ M^{(3)}(s, t) &= \frac{M^{(1)}(s, t) \left(1 - \frac{6B}{B} + \frac{6B^2}{B^2}\right)}{M^{(1)}(s, t) \left(1 - \frac{14B}{B} + \frac{36B^2}{B^2} - \frac{24B^3}{B^3}\right)} \\ M^{(4)}(s, t) &= M^{(1)}(s, t) \left(1 - \frac{14B}{B} + \frac{36B^2}{B^2} - \frac{24B^3}{B^3}\right), \end{aligned}$$

from which we deduce the (non-central) moments $M_k(t) = \mathbb{E}\{Z^k(t)\}$, $k = 1, 2, 3, 4$, by setting $s = 0$.

References

- Allen, LJS. An Introduction to Stochastic Processes with Biology Applications. Prentice Hall; 2003.
- Athreya, KB.; Ney, PE. Branching Processes. Springer-Verlag; Berlin: 1972.
- Becker NG, Hasofer AM. Estimation in epidemics with incomplete observations. J R Statist Soc B. 1997; 59:415–430.
- Boucher K, Yakovlev AY, Mayer-Pröschel M, Noble M. A stochastic model of temporarily regulated generation of oligodendrocytes in vitro. Math Biosci. 1999; 159:47–78. [PubMed: 10361805]
- Boucher K, Zorin A, Yakovlev AY, Mayer-Pröschel M, Noble M. An alternative stochastic model of generation of oligodendrocytes in cell culture. J Math Biol. 2001; 43:22–36. [PubMed: 12120865]
- Carnahan, B.; Luther, HA.; Wilkes, JO. Applied Numerical Methods. John Wiley and Sons; New York: 1969.
- Chen R, Hyrien O, Mayer-Pröschel M, Noble M. A composite likelihood approach to the analysis of longitudinal clonal data under an age-dependent branching process. Biostatistics. 2010 to appear.
- Chow, YS.; Teicher, H. Probability Theory: Independence, Interchangeability, Martingales. Springer; New York: 2003.
- von Collani E, Tsodikov A, Yakovlev A, Mayer-Pröschel M, Noble M. A random walk model of oligodendrocyte generation in vitro and associated estimation problems. Math Biosci. 1999; 159:189. [PubMed: 10414033]
- Concordet D, Nuñez O. A simulated pseudo-maximum likelihood estimator for nonlinear mixed models. Computational Statistics and Data Analysis. 2002; 39:187–201.
- Darwin JH. The behavior of an estimator for a simple birth-and-death process. Biometrika. 1956; 43:23–31.
- Dion, JP.; Keiding, N. Statistical inference in branching processes. In: Joffe, A.; Ney, PE., editors. Branching Processes. Dekker; New York: 1978. p. 105-140.
- Dion JP, Yanev NM. Statistical inference for branching processes with an increasing random number of ancestors. J Statist Planning Inf. 1994; 39:329–359.
- Godambe VP. The foundations of finite sample estimation in stochastic processes. Biometrika. 1985; 72:419–428.
- Gourieroux C, Monfort A, Trognon A. Pseudo maximum likelihood methods: theory. Econometrica. 1984; 52(3):681–700.
- Guttorp, P. Statistical Inference for Branching Processes. John Wiley and Sons; New York: 1991.

- Haccou, P.; Jagers, P.; Vatutin, VA. *Branching Processes: Variation, Growth and Extinction of Populations*. Cambridge University Press; Cambridge: 2005.
- Heyde CC. Extension of a result of Seneta for the supercritical Galton-Watson process. *Ann Math Statist*. 1970; 41:739–742.
- Heyde CC. Remarks on efficiency in estimation for branching processes. *Biometrika*. 1975; 62:49–55.
- Heyde, CC. *Quasi-Likelihood and Its Applications A General Approach to Optimal Parameter Estimation*. Springer-Verlag; New-York.: 1997.
- Hoel DG, Crump KS. Estimating the generation-time distribution of an age-dependent branching process. *Biometrics*. 1974; 30:125–135. [PubMed: 4813378]
- Hyrien O. A pseudo maximum likelihood estimator for discretely observed multi-type Bellman-Harris branching processes. *J Statist Planning Inf*. 2007; 137:1375–1388.
- Hyrien O, Ambeskovic I, Mayer-Pröschel M, Noble M, Yakovlev A. Stochastic Modeling of oligodendrocyte generation in cell culture: model validation with time-lapse data. *Theor Biol Medical Model*. 2006; 3:21.
- Hyrien O, Chen R, Mayer-Pröschel M, Noble M. Saddlepoint approximations to the moments of multi-type age-dependent branching processes, with applications. *Biometrics*. 2010a; 66:567–577. [PubMed: 19508238]
- Hyrien O, Chen R, Zand MS. An age-dependent branching process model for the analysis of CFSE-labeling experiments. *Biology Direct*. 2010b; 5:41. [PubMed: 20569476]
- Hyrien O, Dietrich J, Noble M. Mathematical and experimental approaches to identify and predict the effects of chemotherapy on neuroglial precursors. *Cancer Research*. 2010c to appear.
- Hyrien O, Mayer-Pröschel M, Noble M, Yakovlev A. A stochastic model to analyze clonal data on multi-type cell populations. *Biometrics*. 2005a; 61:199–207. [PubMed: 15737094]
- Hyrien O, Mayer-Pröschel M, Noble M, Yakovlev A. Estimating the life-span of oligodendrocytes from clonal data on their development in cell culture. *Math Biosci*. 2005b; 193:255–274. [PubMed: 15748733]
- Hyrien O, Zand MS. A mixture model with dependent observations for the analysis of CFSE-labeling experiments. *J Am Statist Ass*. 2008; 103:222–239.
- Jacob C. Conditional least squares estimation in nonstationary nonlinear stochastic regression models. *Ann Statist*. 2010; 38:566–597.
- Jagers, P. *Branching Processes with Biological Applications*. John Wiley and Sons; London: 1975.
- Johnson, NI.; Kotz, S. *Continuous Univariate Distributions I*. Houghton Mifflin Company; Boston: 1970.
- Keiding N. Estimation in the birth process. *Biometrika*. 1974; 61:71–80.
- Keiding N. Maximum likelihood Estimation in the birth-and-death process. *Ann Statist*. 1975; 2:363–372.
- Kendall DG. Stochastic processes and population growth. *J R Statist Soc B*. 1949; 11:230–264. with discussion.
- Kimmel, M.; Axelrod, DE. *Branching Processes in Biology*. Springer; New York: 2002.
- Klimko LA, Nelson PI. On conditional least squares estimation for stochastic processes. *Ann Statist*. 1978; 6:629–642.
- Nedelman J, Downs H, Pharr P. Inference on an age-dependent, multitype branching-process model of mast cells. *J Math Biol*. 1985; 25:203–226. [PubMed: 3611982]
- Oh C, Severo NC, Slivka J. Approximations to the maximum likelihood estimate in some pure birth processes. *Biometrika*. 1991; 78:295–299.
- Sevast'yanov, BA. *Branching Processes*. MIR; Moscow: 1971. in Russian
- Sevast'yanov BA, Chistyakov VP. Multidimensional renewal equations and moments of branching processes. *Theor Probab Appl*. 1971; 16:199–214.
- Wedderburn RWM. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*. 1974; 61:439–447.
- Yakovlev AY, Boucher K, Mayer-Pröschel M, Noble M. Quantitative insight into proliferation and differentiation of oligodendrocyte-type 2 astrocyte progenitor cells in vitro. *Proc Natl Acad Sci USA*. 1998a; 95:14164–14167. 4. [PubMed: 9826671]

- Yakovlev AY, Mayer-Pröschel M, Noble M. A stochastic model of brain cell differentiation in tissue culture. *J Math Biol.* 1998b; 37:49–60. [PubMed: 9710974]
- Yakovlev AY, Mayer-Pröschel M, Noble M. Stochastic formulations of a clock model for temporally regulated generation of oligodendrocytes *in vitro*. *Math Comput Model.* 2000; 32:125–137.
- Yakovlev, AY.; Yanev, NM. *Transient Processes in Cell Proliferation Kinetics*. Springer; Heidelberg: 1989.
- Yakovlev AY, Yanev NM. Relative frequencies in multitype branching processes. *Ann Appl Probab.* 2009; 19:1–14.
- Yanev NM. On the statistics of branching processes. *Theory Prob Appl.* 1975; 20:612–622.
- Yanev, NM. Statistical inference for branching processes. In: Ahsanullah, M.; Yanev, GP., editors. *Records and Branching Processes*. Nova Science Publishers Inc.; New York: 2008.
- Zorin AA, Yakovlev AY, Mayer-Pröschel M, Noble M. Estimation problems associated with stochastic modeling of proliferation and differentiation of O-2A progenitor cells *in vitro*. *Math Biosci.* 2000; 167:109–121. [PubMed: 10998484]

Table 1

Means, standard errors (se), and mean squared errors (mse) for three estimators of the mean lifespan for a pure birth process: quasi-likelihood (QLE), Gaussian pseudo-maximum likelihood (PLE), and conditional Gaussian pseudo-maximum likelihood (CPLE) estimators. Also reported is the asymptotic standard error (asympt.) of the QLE, which provided good approximations to the true ones for $n \geq 50$.

n	QLE (= MLE)			PLE			CPLE		
	mean	se (asympt.)	mse	mean	se	mse	mean	se	mse
10	20.06	0.70 (0.61)	0.49	20.34	0.91	0.94	20.17	0.78	0.64
20	20.09	0.49 (0.43)	0.25	20.18	0.60	0.39	20.08	0.53	0.29
50	20.01	0.28 (0.27)	0.08	20.07	0.35	0.13	20.03	0.32	0.10
100	20.01	0.19 (0.19)	0.04	20.03	0.24	0.06	20.02	0.22	0.05
200	20.01	0.15 (0.14)	0.02	20.03	0.18	0.03	20.01	0.15	0.02

True value: $\lambda^* = 1/\theta^* = 20$ hours.

Table 2

Means, standard errors (se), and mean squared errors (mse) for various estimators of the birth and the death rates of a linear birth-and-death process: quasi-likelihood (QLE), conditional Gaussian pseudo-maximum likelihood (CPLE), conditional least square (CLSE), quasi-likelihood conditional on non-extinction (CExQLE), Gaussian pseudo-maximum likelihood combined with MLE for $\delta = \phi - \rho$ (MLE-PLE), average Gaussian pseudo-maximum likelihood (AvPLE), Gaussian pseudo-maximum likelihood (PLE), conditional quasi-likelihood (CQLE), Gaussian pseudo-maximum likelihood conditional on non-extinction (CExPLE), and weighted quasi-likelihood (WQLE) estimators.

true value	n	mean	se	mse [†]	mean	se	mse [†]	mean	se	mse [†]	mean	se	mse [†]
$\phi^* = 0.05/h$	MLE												
	50	0.0496	0.0042	1.80	0.0492	0.0072	5.26	0.0486	0.0077	6.15	0.0501	0.0084	7.08
	100	0.0500	0.0028	0.81	0.0501	0.0048	2.35	0.0496	0.0059	3.51	0.0501	0.0062	3.79
	200	0.0500	0.0020	0.41	0.0501	0.0035	1.19	0.0500	0.0043	1.81	0.0503	0.0042	1.78
	500	0.0500	0.0013	0.17	0.0500	0.0022	0.48	0.0499	0.0027	0.73	0.0500	0.0028	0.76
$\rho^* = 0.025/h$	CPLE												
	5000	0.0500	0.0004	0.01	0.0500	0.0007	0.05	0.0500	0.0008	0.06	0.0500	0.0008	0.07
	CLSE												
	50	0.0249	0.0041	1.71	0.0245	0.0071	5.09	0.0240	0.0076	5.82	0.0254	0.0083	6.85
	100	0.0251	0.0028	0.78	0.0253	0.0048	2.28	0.0248	0.0059	3.45	0.0253	0.0061	3.77
$\phi^* = 0.025/h$	CQLE												
	200	0.0250	0.0019	0.37	0.0251	0.0034	1.16	0.0250	0.0042	1.78	0.0254	0.0041	1.72
	500	0.0250	0.0013	0.16	0.0251	0.0022	0.48	0.0250	0.0027	0.74	0.0250	0.0028	0.76
	5000	0.0250	0.0004	0.01	0.0250	0.0007	0.04	0.0250	0.0008	0.06	0.0250	0.0008	0.07
	MLE-PLE												
	50	0.0249	0.0041	1.71	0.0245	0.0071	5.09	0.0240	0.0076	5.82	0.0254	0.0083	6.85
$\phi^* = 0.05/h$	AvPLE												
	50	0.0490	0.0090	8.14	0.0498	0.0106	11.27	0.0483	0.0096	9.56	0.0453	0.0109	14.16
	100	0.0502	0.0069	4.74	0.0500	0.0076	5.80	0.0492	0.009	8.09	0.0462	0.0078	7.54
	200	0.0501	0.0048	2.28	0.0504	0.0055	3.00	0.0491	0.0085	7.24	0.0462	0.0057	4.65
	500	0.0497	0.0031	0.96	0.0501	0.0034	1.14	0.0491	0.0068	4.68	0.0457	0.0036	3.15
$\rho^* = 0.025/h$	CExPLE												
	5000	0.0500	0.0010	0.09	0.0500	0.001	0.10	0.0493	0.0019	0.43	0.0458	0.0011	1.86
	WQLE												
	50	0.0245	0.0092	8.46	0.0252	0.0113	12.85	0.0236	0.0096	9.39	0.0208	0.0125	17.49
	100	0.0255	0.0072	5.24	0.0252	0.0081	6.52	0.0244	0.0089	7.97	0.0218	0.0091	9.32
$\phi^* = 0.025/h$	WQLE												
	200	0.0251	0.0050	2.47	0.0254	0.0059	3.45	0.0242	0.0085	7.31	0.0215	0.0066	5.62
	500	0.0247	0.0032	1.03	0.0252	0.0037	1.35	0.0241	0.0068	4.64	0.0210	0.0042	3.34
	5000	0.0250	0.0010	0.10	0.0250	0.0011	0.12	0.0243	0.0019	0.42	0.0211	0.0013	1.65
	WQLE												
	5000	0.0250	0.0010	0.10	0.0250	0.0011	0.12	0.0243	0.0019	0.42	0.0211	0.0013	1.65

[†] values multiplied by 10^{-5} .

Table 3

Means, standard errors (se), and mean squared errors (mse) for two estimators of the parameters for a two-type process: quasi-likelihood (QLE) and Gaussian pseudo-maximum likelihood (PLE) estimators.

true value		$\lambda_1^*=58.0$ hours			$\lambda_2^*=24.0$ hours			$p^* = 0.70$		
n	estimator	mean	se	mse	mean	se	mse	mean	se	mse
50	QLE	57.5	4.3	18.7	24.6	9.4	88.7	0.70	0.028	0.0008
	PLE	57.7	4.8	23.1	25.7	10.1	104.9	0.70	0.033	0.0011
100	QLE	57.5	3.4	11.8	24.8	6.9	48.3	0.70	0.021	0.0004
	PLE	58.0	3.8	14.4	25.0	7.7	60.3	0.70	0.025	0.0006
200	QLE	57.7	2.6	5.9	24.6	5.8	34.0	0.70	0.017	0.0003
	PLE	58.0	2.5	6.3	24.5	5.4	29.4	0.70	0.018	0.0003
500	QLE	58.0	1.7	2.9	24.0	3.5	12.3	0.70	0.010	0.0001
	PLE	58.0	1.6	2.6	24.0	3.1	9.6	0.70	0.011	0.0001