University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Davison, AC;Hautphenne, S;Kraus, A

Title:
Parameter estimation for discretely observed linear birth-and-death processes

Date:
2021-03-01

Citation:
Davison, A. C., Hautphenne, S. & Kraus, A. (2021). Parameter estimation for discretely observed linear birth-and-death processes. Biometrics, 77 (1), pp.186-196. https://doi.org/10.1111/biom.13282.

Persistent Link:
https://hdl.handle.net/11343/249545

Parameter estimation for discretely-observed linear birth-and-death processes

A. C. Davison

Ecole Polytechnique Fédérale de Lausanne, EPFL-FSB-MATH-STAT, Station 8, 1015 Lausanne, Switzerland

email: anthony.davison@epfl.ch

and

S. Hautphenne

School of Mathematics and Statistics, The University of Melbourne, 3010 Melbourne, Australia,

and

Ecole Polytechnique Fédérale de Lausanne, EPFL-FSB-MATH-STAT, Station 8, 1015 Lausanne, Switzerland

email: sophiemh@unimelb.edu.au

and

A. Kraus

Department of Mathematics and Statistics, Masaryk University, Brno, Czech Republic.

email: andrea.kraus@mail.muni.cz

SUMMARY: Birth-and-death processes are widely used to model the development of biological populations. Although they are relatively simple models, their parameters can be challenging to estimate, as the likelihood can become numerically unstable when data arise from the most common sampling schemes, such as annual population censuses. A further difficulty arises when the discrete observations are not equi-spaced, e.g., when census data are unavailable for some years. We present two approaches to estimating the birth, death, and growth rates of a discretely-observed linear birth-and-death process: via an embedded Galton–Watson process and by maximizing a saddlepoint approximation

to the likelihood. We study asymptotic properties of the estimators, compare them on numerical examples, and apply the methodology to data on monitored populations.

Key words: Galton–Watson process; Gaussian approximation; Likelihood; Linear birth-and-death process; Saddlepoint approximation.

1. Introduction

Linear birth-and-death processes (LBDPs) are among the simplest and most widely-used continuous-time Markovian population models, with applications in biology, genetics, ecology, and epidemiology (Karlin and Taylor, 1975; Novozhilov et al., 2006). Most estimation methods rely on continuous observation of the population, which allows maximum likelihood estimation of the birth and death rates λ and μ (Keiding, 1975). Continuous measurement is rare in practice, however, and often observations are available only at discrete, perhaps unequally-spaced, times. For instance, data in population biology often arise from annual animal censuses that may be performed at multiple locations, but not every year. This is the case for the black robin *Petroica traversi*, an endangered songbird species endemic to the Chatham Islands, an isolated archipelago located east of New Zealand. By 1980, the population had declined to five birds, including only one successful breeding pair (Elliston, 1994). Through intensive conservation efforts, the population recovered and is now monitored, though without human intervention. The census performed annually on multiple islands of the archipelago during the period of conservation management has since been scaled back in terms of both islands and years covered.

Likelihood-based estimation for discretely-observed LBDPs is difficult, owing to the cumbersome and numerically unstable form of the probability mass function of the population size. Several authors have proposed approaches to address this, and to estimate the parameters of discretely-observed general birth-and-death processes, for which an analytic expression for the population size distribution may be unavailable. Chen and Hyrien (2011) propose quasi- and pseudo-likelihood estimators, but their empirical analysis suggests that these are inferior to the maximum likelihood estimator (MLE). Crawford and Suchard (2012) and Crawford et al. (2014) use continued fractions to obtain expressions for the Laplace transform of the population size distribution. They approach the computation of the likelihood as a

missing data problem, and maximize it using an expectation-maximization algorithm. This approach is also used in Xu et al. (2015), who propose a multitype branching process approximation to birth-death-shift processes, but their method has numerical limitations when the population is very large. In this case, Ross et al. (2006) propose approximating a density-dependent Markov population process by an Ornstein–Uhlenbeck process around a stable equilibrium point of the corresponding deterministic trajectory, which leads to substantial simplifications in the likelihood function for the discretely-observed process. However, this approximation cannot readily be applied with an LBDP, whose only equilibrium point is the (unstable) absorbing state 0.

We propose two ways to estimate the parameters of an LBDP. In Section 3, we use the embedded Galton–Watson process arising when the inter-observation intervals are constant (GW approach). In Section 4, we allow non-constant inter-observation intervals and apply a saddlepoint approximation to the population size distribution (SPA approach). The GW approach is natural, but properties of the resulting estimators have not yet been documented. Keiding (1975) studies this problem but assumes that the numbers of individuals with no offspring are also observed at each generation of the embedded process. His estimators correspond to MLEs, while in our case only the estimator of the growth rate, $\omega = \lambda - \mu$, is an MLE. We derive asymptotic properties of the GW estimators for λ, μ, and ω. These estimators are not the most efficient, but we show that they correspond to the MLEs based on a simplified SPA approach; this unveils an interesting connection between the two approaches, as detailed in Section 5, and motivates GW-type estimators for non-constant inter-observation intervals.

Saddlepoint methods provide accurate approximations in many applications. Daniels (1982) applies them to general birth processes, Aït-Sahalia and Yu (2006) use them to construct closed-form approximations to the transition densities of continuous-time Markov processes,

and Pedeli et al. (2015) use them to estimate high-order integer-valued autoregressive processes. We highlight the gain in numerical stability and computational efficiency when using the saddlepoint technique to approximate the population size distribution in an LBDP compared to evaluating its analytical form, and discuss the approximation error.

Simulations in Section 6 indicate that the relative error between the saddlepoint and exact MLEs quickly becomes negligible as the observed population size increases. The SPA estimators have lower mean square error than the GW estimators, but the two become comparable if one increases the number of discrete observations of a non-extinct population, or its size at the first observation time. In Section 7, we apply our methodology to annual censuses of the black robin population. Other applications are presented in the Supporting Information, which also contains proofs, a discussion of the approximation error, theoretical and computational comparisons between our methodology and other available approaches, and an alternative multivariate approximation.

2. Discretely observed linear birth-and-death processes

An LBDP with birth rate λ and death rate μ is a continuous-time Markovian branching process in which an individual's lifetime is exponentially distributed with parameter μ and reproduction occurs according to a Poisson process with parameter λ. The population size increases from k to $k+1$ at rate $k\lambda$, or decreases from k to $k-1$ at rate $k\mu$. The population growth rate, $\omega = \lambda - \mu$, is also called the *Malthusian parameter*.

The conditional probability mass function of the population size $Z(t)$ at time t, given $Z(0) = a$, is $p_k(t;a) := \mathsf{P}\{Z(t) = k \,|\, Z(0) = a\}$ for $a, k \in \mathbb{Z}^+$. The corresponding probability generating function (PGF) is $f(s,t;a) := \sum_{k\geqslant 0} p_k(t;a)s^k = f(s,t)^a$, where $f(s,t) := f(s,t;1)$ satisfies the backward Kolmogorov equation $\partial f(s,t)/\partial t = \mu - (\lambda+\mu)f(s,t) +$

$\lambda^2 f(s,t)^2$, $f(s,0) = s$, which has the explicit solution

$$f(s,t) = \begin{cases} 1 + \dfrac{(\lambda-\mu)(s-1)}{(\lambda s - \mu)\,\exp\{(\mu-\lambda)t\} - \lambda(s-1)}, & \lambda \neq \mu, \\ 1 + \dfrac{(s-1)}{1-\lambda t(s-1)}, & \lambda = \mu. \end{cases} \tag{1}$$

The distribution of $Z(t)$ conditional on $Z(0) = 1$ corresponds to a modified geometric distribution, $p_k(t;1) = \alpha(t)\,\mathbb{1}_{\{k=0\}} + \{1-\alpha(t)\}\{1-\beta(t)\}\beta(t)^{k-1}\,\mathbb{1}_{\{k>0\}}$, where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function, with $\alpha(t) = \mu\{\exp(\omega t) - 1\}/\{\lambda \exp(\omega t) - \mu\}$ and $\beta(t) = \lambda\alpha(t)/\mu$ if $\lambda \neq \mu$, and $\alpha(t) = \beta(t) = \lambda t/(1+\lambda t)$ if $\lambda = \mu$. Conditionally on $Z(0) = a$, $p_k(t;a) = \alpha(t)^a$ for $k = 0$, and for $k \geqslant 1$,

$$p_k(t;a) = \sum_{j=\max(0,a-k)}^{a-1} \binom{a}{j} \binom{k-1}{a-j-1} \alpha(t)^j \left[\{1-\alpha(t)\}\{1-\beta(t)\}\right]^{a-j} \beta(t)^{k-a+j}. \tag{2}$$

The conditional mean and variance of the population size at time t are (e.g., Guttorp, 1991),

$$m(t;a) = \mathsf{E}\{Z(t) \,|\, Z(0) = a\} = a\,\exp(\omega t), \tag{3}$$

$$\sigma^2(t;a) = \mathsf{Var}\{Z(t) \,|\, Z(0) = a\} = \begin{cases} a\,\frac{\lambda+\mu}{\omega}\,\exp(\omega t)\{\exp(\omega t) - 1\}, & \lambda \neq \mu, \\ a\,2\lambda t, & \lambda = \mu. \end{cases} \tag{4}$$

We write $m(t) := m(t;1)$ and $\sigma^2(t) := \sigma^2(t;1)$.

Suppose that M independent trajectories of an LBDP are each discretely observed. Suppose that trajectory i is observed on $N_i + 1$ separate occasions, where $N_i \geqslant 1$, let $t_{i,j}$ and $k_{i,j}$ $(j = 0, \ldots, N_i)$ denote the time of the $(j+1)$st observation and the corresponding population size, where $t_{i,0} = 0$, and let $\tau_{i,j} := t_{i,j} - t_{i,j-1}$ $(j = 1, \ldots, N_i)$ denote the jth inter-observation interval. The log-likelihood based on observations $\boldsymbol{t} := \{t_{i,j}\}$ and $\boldsymbol{k} := \{k_{i,j}\}$ is

$$\ell(\lambda, \mu; \boldsymbol{t}, \boldsymbol{k}) = \sum_{i=1}^{M} \log \mathsf{P}\{Z(t_{i,1}) = k_{i,1}, \ldots, Z(t_{i,N_i}) = k_{i,N_i} \,|\, Z(0) = k_{i,0}\} \tag{5}$$

$$= \sum_{i=1}^{M} \sum_{j=1}^{N_i} \log \mathsf{P}\{Z(\tau_{i,j}) = k_{i,j} \,|\, Z(0) = k_{i,j-1}\} = \sum_{i=1}^{M} \sum_{j=1}^{N_i} \log p_{k_{i,j}}(\tau_{i,j}; k_{i,j-1}). \tag{6}$$

Computing the MLEs therefore requires the efficient calculation of $p_k(t;a)$ for any $a, k \in \mathbb{Z}^+$ and $t \in \mathbb{R}^+$. Despite the explicit form (2) of these probabilities, the binomial coefficients

increase rapidly with k and a, leading to numerical difficulties, and making the maximization of (6) cumbersome, if not impossible. These issues, mentioned earlier in the literature (for example in Darwin, 1956 or Chen and Hyrien, 2011), hamper the use of likelihood estimation of λ and μ from discrete observations. In the next two sections we propose methods to circumvent these difficulties, and then we connect them in Section 5.

3. Embedded Galton–Watson process approach

Suppose we observe a single trajectory of the LBDP, $M = 1$, and the inter-observation intervals τ_j all equal τ. The discretely-observed LBDP then corresponds to a Galton–Watson process with offspring PGF $P(s) := f(s, \tau)$ and offspring mean and variance $m := m(\tau)$ and $\sigma^2 := \sigma^2(\tau)$ (Harris, 2002). The birth and death rates λ and μ are related to m and σ^2 via

$$\lambda = \frac{\log m}{2\tau}\left\{\frac{\sigma^2}{m(m-1)} + 1\right\}, \quad \mu = \frac{\log m}{2\tau}\left\{\frac{\sigma^2}{m(m-1)} - 1\right\} \tag{7}$$

if $m \neq 1$, and $\lambda = \mu = \sigma^2/(2\tau)$ if $m = 1$, so they can be estimated using estimators of m and σ^2, which we now discuss.

Based on observing the initial population size $Z_0 := Z(0)$ and the successive population sizes $Z_1 := Z(\tau), \ldots, Z_N := Z(N\tau)$, m and σ^2 can be estimated by

$$\widehat{m}_{Z_0,N} = \frac{\sum_{j=1}^{N} Z_j}{\sum_{j=1}^{N} Z_{j-1}}, \qquad \widehat{\sigma^2}_{Z_0,N} = \frac{1}{N}\sum_{j=1}^{N} Z_{j-1}\left(\frac{Z_j}{Z_{j-1}} - \widehat{m}_{Z_0,N}\right)^2, \tag{8}$$

where we set $Z_j/Z_{j-1} := 1$ when $Z_{j-1} = 0$ (implying that $Z_j = 0$); $\widehat{m}_{Z_0,N}$ is a method of moments, conditional least squares, and non-parametric maximum likelihood estimator of the offspring mean (Guttorp, 1991).

Asymptotic properties of $\widehat{m}_{Z_0,N}$ and $\widehat{\sigma^2}_{Z_0,N}$ have been studied extensively. As $N \to \infty$, and conditionally on the survival of the process, $\{\lim_{j\to\infty} Z_j = \infty\}$, both are consistent and

asymptotically normal (Guttorp, 1991):

$$\sqrt{\frac{\sum_{j=1}^{N} Z_{j-1}}{\sigma^2}}\left(\widehat{m}_{Z_0,N} - m\right) \xrightarrow{d} \mathcal{N}(0,1), \tag{9}$$

$$\sqrt{\frac{N}{2\sigma^4}}\left(\widehat{\sigma^2}_{Z_0,N} - \sigma^2\right) \xrightarrow{d} \mathcal{N}(0,1), \tag{10}$$

where $\xrightarrow{d}$ denotes convergence in distribution. These results hold for any fixed value of Z_0, but only in the supercritical case, $m > 1$, in which survival occurs with positive probability. If both $Z_0 \to \infty$ and $N \to \infty$, and conditionally on $\{\lim_{j\to\infty} Z_j = \infty\}$, the estimators are consistent and jointly asymptotically normal (Duby and Rouault, 1982):

$$\begin{pmatrix} \sqrt{\frac{Z_0\,(m^N-1)}{\sigma^2\,(m-1)}}\,(\widehat{m}_{Z_0,N} - m) \\ \sqrt{\frac{N}{2\sigma^4}}(\widehat{\sigma^2}_{Z_0,N} - \sigma^2) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{I}\right), \tag{11}$$

where $\mathcal{N}(\mathbf{0}, \mathbf{I})$ is the bivariate normal distribution with independent standard normal marginals. Properties of $\widehat{m}_{Z_0,N}$ and $\widehat{\sigma^2}_{Z_0,N}$ under other asymptotic regimes and for critical or subcritical processes, $m = 1$ or $m < 1$, have also been studied (Guttorp, 1991; Dion and Yanev, 1994), but we shall not develop them here.

On inserting the estimators from (8) into (7), we obtain estimators for λ and μ, i.e.,

$$\widehat{\lambda}_{Z_0,N} = \frac{\log \widehat{m}_{Z_0,N}}{2\tau}\left\{\frac{\widehat{\sigma^2}_{Z_0,N}}{\widehat{m}_{Z_0,N}(\widehat{m}_{Z_0,N}-1)} + 1\right\}, \tag{12}$$

$$\widehat{\mu}_{Z_0,N} = \frac{\log \widehat{m}_{Z_0,N}}{2\tau}\left\{\frac{\widehat{\sigma^2}_{Z_0,N}}{\widehat{m}_{Z_0,N}(\widehat{m}_{Z_0,N}-1)} - 1\right\}. \tag{13}$$

Their properties, summarized in the next theorem, can be derived from those of $\widehat{m}_{Z_0,N}$ and $\widehat{\sigma^2}_{Z_0,N}$ (see Web Appendix A for a proof).

THEOREM 1: *Suppose that* $m > 1$. *Then* $\widehat{\lambda}_{Z_0,N}$ *and* $\widehat{\mu}_{Z_0,N}$ *are consistent on* $\{\lim_{j\to\infty} Z_j = \infty\}$ *as* $N \to \infty$. *Moreover, conditionally on* $\{\lim_{j\to\infty} Z_j = \infty\}$,

$$\sqrt{N}\begin{pmatrix} \widehat{\lambda}_{Z_0,N} - \lambda \\ \widehat{\mu}_{Z_0,N} - \mu \end{pmatrix} \xrightarrow{d} \mathcal{N}\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{(\log m)^2\sigma^4}{2\tau^2 m^2(m-1)^2} \times \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right\}$$

as $N \to \infty$. *The same statements hold as both* $N \to \infty$ *and* $Z_0 \to \infty$.

By Theorem 1, the difference $\sqrt{N}\left\{(\widehat{\lambda}_{Z_0,N} - \lambda) - (\widehat{\mu}_{Z_0,N} - \mu)\right\}$ vanishes in probability conditional on $\{\lim_{j\to\infty} Z_j = \infty\}$ as $N \to \infty$, and also as both $N \to \infty$ and $Z_0 \to \infty$. Thus, the estimator $\widehat{\omega}_{Z_0,N} = \widehat{\lambda}_{Z_0,N} - \widehat{\mu}_{Z_0,N}$ of ω satisfies $\sqrt{N}(\widehat{\omega}_{Z_0,N} - \omega) \xrightarrow{p} 0$ conditional on $\{\lim_{j\to\infty} Z_j = \infty\}$ as $N \to \infty$, and as both $N \to \infty$ and $Z_0 \to \infty$, where $\xrightarrow{p}$ denotes convergence in probability. In fact, as $\widehat{\omega}_{Z_0,N} = \log(\widehat{m}_{Z_0,N})/\tau$ depends only on $\widehat{m}_{Z_0,N}$, its asymptotic behavior, including the faster rate of convergence than for $\widehat{\lambda}_{Z_0,N}$ and $\widehat{\mu}_{Z_0,N}$, is determined by the asymptotic behavior of $\widehat{m}_{Z_0,N}$. This is made precise in the next theorem, proved in Web Appendix A.

THEOREM 2: *Suppose that $m > 1$. Then $\widehat{\omega}_{Z_0,N}$ is consistent on $\{\lim_{j\to\infty} Z_j = \infty\}$ as $N \to \infty$, and also as both $N \to \infty$ and $Z_0 \to \infty$. Moreover, conditionally on $\{\lim_{j\to\infty} Z_j = \infty\}$,*

$$\sqrt{\sum_{j=1}^{N} Z_{j-1}} \times \left(\widehat{\omega}_{Z_0,N} - \omega\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2}{m^2\tau^2}\right), \qquad \text{as } N \to \infty,$$

$$\sqrt{Z_0\left(m^N - 1\right)} \times \left(\widehat{\omega}_{Z_0,N} - \omega\right) \xrightarrow{d} \mathcal{N}\left\{0, \frac{\sigma^2(m-1)}{m^2\tau^2}\right\}, \qquad \text{as } N \to \infty, Z_0 \to \infty.$$

Suppose now that we observe $M > 1$ independent trajectories of the LBDP, and the inter-observation intervals $\tau_{i,j}$ all equal τ. Let $Z_{i,j} := Z_i(j\tau)$ be the $(j+1)$st observation of the ith trajectory ($i = 1, \ldots, M,\ j = 0, \ldots, N_i$). Then the estimators (8) generalize to

$$\widehat{m}_{M,\mathbf{N}} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N_i} Z_{i,j}}{\sum_{i=1}^{M}\sum_{j=1}^{N_i} Z_{i,j-1}}, \qquad \widehat{\sigma^2}_{M,\mathbf{N}} = \frac{1}{\sum_{i=1}^{M} N_i}\sum_{i=1}^{M}\sum_{j=1}^{N_i} Z_{i,j-1}\left(\frac{Z_{i,j}}{Z_{i,j-1}} - \widehat{m}_{M,\mathbf{N}}\right)^2, \quad (14)$$

where we again set $0/0 := 1$. These estimators appear not to have been proposed previously. We derive them in Section 5 (see Lemma 2 and its proof), where we also establish the consistency and asymptotic normality of the corresponding estimators of λ and μ (Theorem 3).

4. Saddlepoint approximation

Saddlepoint methods (Daniels, 1954; Butler, 2007) offer highly accurate approximations for probability densities and distributions that are particularly useful when these are intractable or when their numerical evaluation is unstable. Let Y be a random variable on $\mathbb{R}^+$ with density $h(y)$ and cumulant generating function (CGF) $K(x)$ and suppose that we wish to approximate the density $h_n(y)$ of the sum of n independent copies of Y. We start by writing

$$h(y) = \exp\{K(x) - yx\}\, h(y; x),$$

where $h(y; x) := \exp\{yx - K(x)\}\, h(y)$ is a tilted version of $h(y)$. Then a sum Y_n of n independent variables from the tilted density has density $h_n(y; x) = \exp\{yx - nK(x)\}\, h_n(y)$, and Y_n has CGF $K_n(x^*) = n\{K(x^* + x) - K(x)\}$, so its mean and variance are $nK'(x)$ and $nK''(x)$. In order to approximate $h_n(y)$, we choose x equal to the value $\tilde{x}$ for which $nK'(\tilde{x}) = y$, so the corresponding tilted density $h_n(y; \tilde{x})$ has mean y, and replace $h_n(y; \tilde{x})$ by a normal approximation evaluated at its mean, yielding

$$\begin{aligned} h_n(y) &= \exp\{nK(\tilde{x}) - y\tilde{x}\}\, h_n(y; \tilde{x}) \\ &\doteq \exp\{nK(\tilde{x}) - y\tilde{x}\}\, \frac{1}{\sqrt{2\pi K''(\tilde{x})}}. \end{aligned} \tag{15}$$

The error of the *saddlepoint approximation* (15) is $O(n^{-1})$ owing to the symmetry of the normal approximation. We now apply it to the population size at time t in an LBDP.

The conditional CGF of the population size $Z(t)$ at time t, given $Z(0) = a$, is defined as $K(x, t; a) := \log M(x, t; a)$, where $M(x, t; a) := f\{\exp(x), t; a\}$ is the conditional MGF. For a given t, the maximal convergence set of $M(x, t; a)$ in a neighborhood of 0 is $\mathcal{S}(t) = (-\infty, \log R(t))$, where $R(t) = 1+\omega/[\lambda\{\exp(\omega t)-1\}]$ if $\lambda \neq \mu$, and $R(t) = 1+1/(\lambda t)$ if $\lambda = \mu$. When $k = 0$, we have $p_k(t; a) = \alpha(t)^a$ and no approximation is needed. For $k = 1, 2, \ldots$, the saddlepoint approximation of $p_k(t; a)$ is (see Butler, 2007, for instance)

$$\tilde{p}_k(t; a) = \frac{1}{\sqrt{2\pi K''(\tilde{x}, t; a)}} \exp\{K(\tilde{x}, t; a) - \tilde{x}k\}, \tag{16}$$

where the saddlepoint $\tilde{x} = \tilde{x}(k,t;a)$ corresponds to the unique solution in $\mathcal{S}(t)$ to the equation

$$K'(\tilde{x},t;a) := \frac{\partial}{\partial x}K(x,t;a)\Big|_{x=\tilde{x}} = k. \tag{17}$$

Since $K(x,t;a) = aK(x,t;1)$, (16) can be rewritten as

$$\tilde{p}_k(t;a) = \frac{1}{\sqrt{2\pi a K''(\tilde{x},t;1)}}\exp[a\{K(\tilde{x},t;1) - (\tilde{x}\,k/a)\}], \tag{18}$$

where $\tilde{x} = \tilde{x}(k/a,t;1)$ is the unique solution in $\mathcal{S}(t)$ to $K'(\tilde{x},t;1) = k/a$.

The next lemma gives analytical expressions for $\tilde{x}(k,t;a)$ and $\tilde{p}_k(t;a)$ in the LBDP case.

LEMMA 1: *The saddlepoint and saddlepoint approximation for an LBDP can be expressed as* $\tilde{x}(k,t;a) = \log\tilde{s}$, *where* $\tilde{s} := \tilde{s}(k,t;a) = (2A)^{-1}(-B + \sqrt{B^2 - 4AC})$, *with, for* $\lambda \neq \mu$,

$$\begin{aligned} A &:= A(t) &&= \lambda\{m(t)-1\}\{\lambda - \mu m(t)\}, \\ B &:= B(k,t;a) &&= 2\lambda\mu\{1 + m(t)^2 - m(t) - (a/k)m(t)\} \\ &&&\quad + m(t)(\lambda^2+\mu^2)\{(a/k)-1\}, \\ C &:= C(t) &&= \mu\{m(t)-1\}\{\mu - \lambda m(t)\}, \end{aligned}$$

leading to

$$\begin{aligned} \tilde{p}_k(t;a) &= \frac{1}{\sqrt{2\pi a}}\frac{1}{\tilde{s}^k}\left\{\frac{\mu - \lambda\tilde{s} + \mu(\tilde{s}-1)\,m(t)}{\mu - \lambda\tilde{s} + \lambda(\tilde{s}-1)\,m(t)}\right\}^a \times \\ &\quad \left\{-\frac{\{m(t)-1\}m(t)\tilde{s}(\lambda-\mu)^2\,\{-\lambda^2\tilde{s}^2 + \lambda m(t)\mu\,(\tilde{s}^2-1) + \mu^2\}}{\{\lambda[m(t)(\tilde{s}-1) - \tilde{s}] + \mu\}^2\{\lambda\tilde{s} + \mu[-m(t)\tilde{s} + m(t) - 1]\}^2}\right\}^{-1/2}, \end{aligned} \tag{19}$$

and for $\lambda = \mu$, $A := A(t) = \lambda t - (\lambda t)^2$, $B := B(k,t;a) = 2(\lambda t)^2 + (a/k) - 1$, $C := C(t) = -\lambda t - (\lambda t)^2$, *leading to*

$$\tilde{p}_k(t;a) = \frac{1}{\sqrt{2\pi a}}\frac{1}{\tilde{s}^k}\left\{\frac{\lambda t(1-\tilde{s}) + \tilde{s}}{1 - \lambda t(\tilde{s}-1)}\right\}^a\left\{\frac{\lambda t\tilde{s}\,(-\lambda t\tilde{s}^2 + \lambda t + \tilde{s}^2 + 1)}{\{\lambda t(\tilde{s}-1) - 1\}^2(-\lambda t\tilde{s} + \lambda t + \tilde{s})^2}\right\}^{-1/2}.$$

The approximations $\tilde{p}_k(t;a)$ may not sum to unity over k, so it might appear preferable to use the *normalized saddlepoint approximation* (Butler, 2007, p.9),

$$\bar{p}_0(t;a) = p_0(t;a), \quad \bar{p}_k(t;a) = \{1 - p_0(t;a)\}\tilde{p}_k(t;a)\left\{\sum_{j\geqslant 1}\tilde{p}_j(t;a)\right\}^{-1}, \quad k \geqslant 1.$$

Daniels (1982) points out that renormalization complicates likelihood estimation, however, so we shall use the unnormalized approximation throughout.

[Figure 1 about here.]

Figure 1 illustrates how the ratios of the normalized and unnormalized saddlepoint approximations to the exact distribution converge to unity as the initial population size a increases: the approximation improves when the numerical computation of the exact distribution becomes more challenging. The gain in CPU time when computing the saddlepoint approximation rather than the exact distribution is illustrated in Web Tables 4–7 in Web Appendix E, in which various computational approaches are compared. The saddlepoint approach provides the fastest and most numerically stable approximation.

[Figure 2 about here.]

Figure 2(a) shows that when there is a large gap between $p_0(t;a)$ and $p_1(t;a)$, the saddlepoint approximation $\tilde{p}_k(t;a)$ has a smoothing effect for small integers $k > 0$. This effect decreases as $|p_0(t;a) - p_1(t;a)|$ decreases, i.e., as the probability of extinction by time t decreases. It is possible to adjust the smoothing by applying the saddlepoint technique to the conditional CGF, given that the population is not extinct at time t, i.e.,

$$K\{x,t;a \,|\, Z(t)>0\} = \log\{M(x,t;a) - p_0(t;a)\} - \log\{1-p_0(t;a)\},$$

though the resulting saddlepoint $\tilde{x}(k,t;a)$ must be obtained numerically. This approximation applies only when k is interior to the support of the random variable $Z(t)$ conditional on $Z(t) > 0$, i.e., for $k \geqslant 2$; we use the exact probability $p_1(t;a) = \{\partial f(s,t;a)/\partial s\}|_{s=0}$ for $k = 1$. Figure 2(b) shows the resulting *adjusted (conditional) saddlepoint approximation*, which is unnecessary when k or a exceed 20 or so. This might be implemented even more quickly using the exact derivative methods of Zhang et al. (2019), but we have not investigated this.

The saddlepoint approximated log-likelihood is

$$\tilde{\ell}(\lambda, \mu; \boldsymbol{t}, \boldsymbol{k}) = \sum_{i=1}^{M} \sum_{j=1}^{N_i} \log \tilde{p}_{k_{i,j}}(\tau_{i,j}; k_{i,j-1}), \tag{20}$$

and the values $\tilde{\lambda}$ and $\tilde{\mu}$ maximizing this expression are called the *saddlepoint maximum likelihood estimators* (SPMLEs), or *adjusted* SPMLEs, if they are computed using the adjusted saddlepoint approximation.

As indicated by Figure 1, the approximation error of each individual term in (20) decreases as the population size $k_{i,j-1}$ increases. More precisely (e.g., Davison, 2003, Section 12.3.2),

$$p_k(t; a) = \tilde{p}_k(t; a)\{1 + O(1/a)\}, \quad k = 1, 2, \ldots, \tag{21}$$

i.e., the approximation error decreases to zero as $a \to \infty$. As computation of the exact conditional probabilities $p_k(t; a)$ is tractable for small a, the error is smallest exactly when an approximation for $p_k(t; a)$ is most needed.

The error of the saddlepoint-approximated log-likelihood for the ith trajectory is the sum of the individual errors $\log \tilde{p}_{k_{i,j}}(\tau_{i,j}; k_{i,j-1}) - \log p_{k_{i,j}}(\tau_{i,j}; k_{i,j-1})$ $(j = 1, \ldots, N_i)$, but our experience is that these do not appear to accumulate badly. One reason for this may be the extinction/explosion dichotomy of the LBDP, which suggests that when the trajectory is observed for a long period, the numbers $k_{i,j-1}$ in the conditional probabilities $p_{k_{i,j}}(\tau_{i,j}; k_{i,j-1})$ either remain very small, and thus the conditional probabilities need not be approximated, or become very large, and then the conditional probabilities are approximated very well. Web Appendix B discusses the approximation error in more detail.

An alternative approach avoiding the accumulation of individual errors is to use saddlepoint approximation directly on $\mathsf{P}\{Z(t_{i,1}) = k_{i,1}, \ldots, Z(t_{i,N_i}) = k_{i,N_i} \,|\, Z(0) = k_{i,0}\}$ in (5), rather than passing to (6) and approximating $p_{k_{i,j}}(\tau_{i,j}; k_{i,j-1})$ $(j = 1, \ldots, N_i)$. This *multivariate saddlepoint approximation* is discussed in Web Appendix C. Simulations show no clear

approximation error advantage for either of these approaches. A deeper study of the multivariate approach is an avenue for further work.

The discussion above concerns the error in the saddlepoint-approximated log-likelihood. We explore the error in the resulting estimators, the SPMLEs, by simulation in Section 6. We show first how Gaussian approximation links the Galton–Watson and saddlepoint approaches.

5. Gaussian approximations

The saddlepoint approximation can be simplified by replacing the CGF $K(x,t;a)$ in (16) and (17) by its second-order Taylor expansion at $x=0$,

$$K(x,t;a) \doteq \tilde{K}(x,t;a) := xK'(0,t;a) + \frac{x^2}{2}K''(0,t;a) = xm(t;a) + \frac{x^2}{2}\sigma^2(t;a). \tag{22}$$

Since $\tilde{K}(x,t;a)$ is the CGF of a normal distribution, (22) amounts to a Gaussian approximation to the distribution of $Z(t)$ conditional on $Z(0)=a$. Saddlepoint approximation of a normal distribution is exact (Butler, 2007), so the simplified saddlepoint approximation is

$$\tilde{\varphi}_k(t;a) := \left\{\frac{1}{2\pi\sigma^2(t;a)}\right\}^{1/2} \exp\left[-\frac{\{k-m(t;a)\}^2}{2\sigma^2(t;a)}\right]. \tag{23}$$

The following lemma, proved in Web Appendix A, states that maximizing the saddlepoint approximated log-likelihood (20) in which the saddlepoint approximations $\tilde{p}_{k_{i,j}}(\tau_{i,j};k_{i,j-1})$ are replaced by their simplified versions $\tilde{\varphi}_{k_{i,j}}(\tau_{i,j};k_{i,j-1})$ leads to the GW estimators.

LEMMA 2: *If the inter-observation intervals are constant, $\tau_{i,j}=\tau$ for $i=1,\ldots,M$ and $j=1,\ldots,N_i$, then the MLEs for λ and μ resulting from replacing* (16) *by* (23) *in* (20) *coincide with the GW estimators.*

This connects the GW and SPA approaches, and suggests a natural extension of the GW estimators when the inter-observation intervals are unequal. Using (3) and (4) and letting

$\omega = \lambda - \mu$ and $\xi = \lambda + \mu$, (23) is rewritten as

$$\log \tilde{\varphi}_k(t; a \,|\, \omega, \xi) = -\frac{1}{2} \log \left[2\pi a \, \frac{\xi}{\omega} \exp(\omega t) \{\exp(\omega t) - 1\} \right] - \frac{\omega \{k - a \exp(\omega t)\}^2}{2a\xi \exp(\omega t) \{\exp(\omega t) - 1\}}. \quad (24)$$

Replacing (16) by (24) in (20) gives the simplified saddlepoint approximated log-likelihood

$$\begin{aligned} \tilde{\tilde{\ell}}(\omega, \xi; \boldsymbol{t}, \boldsymbol{k}) \;\; = \;\; & -\frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{N_i} \log \left[2\pi k_{i,j-1} \, \frac{\xi}{\omega} \exp(\omega \tau_{i,j}) \{\exp(\omega \tau_{i,j}) - 1\} \right] \\ & - \frac{\omega}{2\xi} \sum_{i=1}^{M} \sum_{j=1}^{N_i} \frac{\{k_{i,j} - k_{i,j-1} \exp(\omega \tau_{i,j})\}^2}{k_{i,j-1} \exp(\omega \tau_{i,j}) \{\exp(\omega \tau_{i,j}) - 1\}}. \end{aligned} \quad (25)$$

The generalized GW estimators, implicitly defined as maximizers of (25) with respect to λ and μ, enjoy standard asymptotic properties as $M \to \infty$, specified by the following theorem proved in Web Appendix A.

THEOREM 3: *Suppose that $\lambda > 0$, that $0 < \bar{\tau} < \inf_{1 \leqslant j \leqslant N_i; 1 \leqslant i \leqslant M} \tau_{i,j}$, and that observation of each trajectory stops at the earlier of its extinction time T_i and after $N_i + 1$ observations. Then the estimators resulting from maximization of* (25) *are consistent and asymptotically normal as $M \to \infty$, with covariance matrix given by (S.I. 11) in Web Appendix A.*

The approximation (22) corresponds to the distribution at time t of the Gaussian diffusion approximation of the LBDP as $Z(0) \to \infty$. Indeed, if $Z(0) = a$, then the process $\left\{ W^{(a)}(s) := \sqrt{a} \{ Z(s)/a - m(s) \} \right\}_{0 \leqslant s \leqslant t}$ converges weakly to the Gaussian diffusion $W(s)$ with zero mean and with $\mathsf{Cov}\{W(s), W(u)\} = \omega^{-1}(\lambda + \mu) \exp(\omega u) \{\exp(\omega s) - 1\}$ if $\omega \neq 0$, and $\mathsf{Cov}\{W(s), W(u)\} = (\lambda + \mu) s$ if $\omega = 0$, for $0 \leqslant s \leqslant u \leqslant t$ (Barbour, 1974). However, instead of using the approximating diffusion process, here we first split the trajectory-wise likelihoods in (5) by using the Markov property of the LBDP, leading to (6), apply the diffusion approximation separately on each interval $[t_{i,j}, t_{i,j+1}]$, and only use the distributions of the approximating processes at the end of each interval in (25). In the context of Galton–Watson processes, this can be justified by the asymptotic independence of properly standardized

increments of the process (cf. Kvitkovičová and Panaretos, 2011). We explore the multivariate case where (5) is used without (6) in Web Appendix C.

Gaussian approximation only uses the first two moments of the population size, whereas the saddlepoint approximation retains more information. As we shall now see, this distinction affects the quality of the corresponding estimators.

6. Simulations

To compare the SPMLEs with the MLEs, we simulated independent non-extinct trajectories of LBDPs with $\lambda = 7$ and $\mu = 5$, with various initial population sizes. For each trajectory, we recorded the population sizes at $N + 1 = 20$ time points, and used them to calculate the SPMLEs and MLEs. In about a fifth of the trajectories the observed population size exceeded 600 individuals and the MLEs could not be computed.

For the estimators of λ and μ, the relative error between SPMLEs and MLEs is always below 5%, and only exceeds 2% for minimum population sizes smaller than five. For ω, the relative error is generally smaller than 0.01%, so even the saddlepoint approximation errors in the SPMLEs of λ and μ appear to cancel in the SPMLE of ω. The maximum observed population size has no impact on the CPU time necessary to compute the SPMLEs, which was around 0.1s in each case, whereas the CPU time to obtain the exact MLEs increased roughly linearly with the population size at rate around 0.2s/individual.

We ran a series of simulation experiments, increasing either the number N of discrete-time observations or the number M of independent observed trajectories. We first simulated 100 replicates of a single ($M = 1$) non-extinct trajectory of an LBDP with $\lambda = 7$ and $\mu = 6$, starting with $Z(0) = 10$ individuals, and we computed the absolute bias and root mean square error (RMSE) of $\widehat{\lambda}$ and $\widehat{\omega}$ for $N = 10, \ldots, 60$. Figure 3(a) depicts the results for $\widehat{\lambda}$. As N increases, the absolute value of both biases decrease rapidly, with the SPMLE having

smaller bias than the GW estimator, as anticipated at the end of the previous section, and the difference between the RMSEs of the two estimators (SPMLE vs. GW) approaches zero.

[Figure 3 about here.]

In a second experiment, we fixed $N = 30$ and analyzed properties of the estimators for $M = 1, \ldots, 150$, for an LBDP with $\lambda = 7$ and $\mu = 5$ starting with different initial population sizes $Z(0) = Z_0 = 1, 10, 20$. The bias and RMSE of $\widehat{\lambda}$ are presented in Figure 3(b). The quality of the estimators improves rapidly as M increases. The value of Z_0 has a clear impact on the quality of the GW estimators, and less impact on that of the SPMLEs. We further investigated the effect of Z_0 on the RMSE of both estimators for different values of N, M, and the model parameters. The results, summarized in Table 1, confirm the key role played by Z_0 in the quality of the GW estimators, which become comparable to the SPMLEs as Z_0 increases. For the same values of M, N and Z_0, the quality of the estimators appears to improve as the process moves further into the supercritical regime, as suggested theoretically when m increases away from 1 in Theorem 1.

[Table 1 about here.]

7. Application

We use our methodology to estimate the birth and death rates of two endangered bird populations, and the transmission and removal rates for an influenza epidemic. Due to relatively large population sizes, obtaining the exact MLEs was problematic in all three cases, owing to slow convergence with no guarantee of accurate results. Here we present the analysis for the black robin population, and refer to Web Appendix D for the other analyses.

[Figure 4 about here.]

The black robin, mentioned in Section 1, was subject to conservation efforts in the 1980s, which saw the population on the Chatham Islands recover from 5 to 93 birds (Massaro et al., 2013a). In the 1990s, the population was monitored without intervention, and grew to 197 adults by 1998 (Kennedy, 2009). After this period, its growth slowed, reaching 239 adults in 2011 and 298 in 2014 on a single island (Massaro et al., 2013b, Massaro, unpublished data).

We fit an LBDP to the yearly censuses of the female black robin population between 1972 and 1998 ($N = 26$), combined over several islands of the archipelago. The data are presented in Figure 4(a), and the estimates are shown in Table 2(a). The computation of the GW estimates is approximately 160 times faster than the SPMLEs, whose computation is itself 1500 times faster than the exact MLEs. The GW estimates are very close to the MLEs, but the SPMLEs are a little different, which we attribute to small population sizes; the adjusted SPMLEs show a substantial improvement. The table also highlights the time-efficiency of the GW and SPA approaches to computing the standard errors, compared to the exact approach.

[Table 2 about here.]

The SPMLE estimates imply that, during 1972–1998, a one-year-old female bird had a life expectancy of $1/0.17 = 5.78$ (SE 2.40) years (MLE estimates imply $1/0.19 = 5.26$ years, SE 1.95), and gave birth on average to $0.301/0.173 = 1.74$ (SE 0.36) female offspring reaching an age of one year (MLE $0.318/0.19 = 1.67$, SE 0.31); these are biologically reasonable estimates. The estimated probability of extinction for the black robin is $0.173/0.301 = 0.57$ (SE 0.12), compared to $0.19/0.318 = 0.6$ (SE 0.11) using the MLE.

We fit another LBDP to the censuses of the female black robin population on Rangatira Island from 1989 to 2014. The data are shown in Figure 4(b): the gap between 1999 and 2009 was due to the lack of government funding for detailed monitoring of the population (Massaro et al., 2013b, 2018). The estimates are shown in Table 2(b). Compared to the

estimates for the period before 1999, the GW estimates, SPMLEs, and MLEs for the more recent period are more similar. The population sizes for the more recent period are larger than before, so the SPMLE has a smaller overall error and there is less advantage in using its adjusted version. The gain from using SPMLEs over generalized GW estimates is also lower. The values of the estimates indicate a difference in the growth of the populations in the two periods, with higher birth and lower death rates before 1999, perhaps because the fertility rates were increased artificially when the birds were closely monitored in 1980–1989, and competition for resources was lower in the smaller population before 1999.

8. Discussion

We investigated two approaches to the estimation of the birth and death rates of a discretely-observed linear birth and death process (LBDP). The Galton–Watson approach provides simple explicit estimators, and the saddlepoint approach provides an explicit approximation to the log-likelihood. In view of the advantages of the second, we regard the first as being of more theoretical than practical interest, but both are numerically appealing and easily implemented. User-friendly MATLAB code is provided as supplementary material.

As shown in Sections 6, 7, and Web Appendix E, the saddlepoint approach has competitive features: our experiments suggest that it is faster and more stable than well-established methods for inverting probability generating functions for LBDPs and recently-proposed methods for estimating their parameters, and has similar accuracy for populations larger than 20–30 individuals. Moreover it provides standard errors much faster than do its competitors.

Optimization of a log-likelihood function generally benefits from a good initial value, and in our context this is provided by the explicit Galton–Watson estimates, which have guaranteed convergence to the true parameter value.

Acknowledgements

SH, AK and ACD respectively thank the Australian Research Council for support through Discovery Early Career Researcher Award DE150101044, the Czech Science Foundation for support through Grant GJ17-22950Y, and the Swiss National Science Foundation. The authors also thank Melanie Massaro for sharing her expertise on the black robin population, Simon Tavaré and Phil Pollett for fruitful discussions, and the exceptionally helpful reviewers.

Data Availability Statement

The data that support the findings in this paper are available in the Supporting Information.

References

Aït-Sahalia, Y. and Yu, J. (2006). Saddlepoint approximations for continuous-time Markov processes. *Journal of Economics* **134,** 507–551.

Barbour, A. D. (1974). On a functional central limit theorem for Markov population processes. *Advances in Applied Probability* **6,** 21–39.

Butler, R. W. (2007). *Saddlepoint Approximations with Applications.* Cambridge University Press, Cambridge.

Chen, R. and Hyrien, O. (2011). Quasi-and pseudo-maximum likelihood estimators for discretely observed continuous-time Markov branching processes. *Journal of Statistical Planning and Inference* **141,** 2209–2227.

Crawford, F. and Suchard, M. (2012). Transition probabilities for general birth–death processes with applications in ecology, genetics, and evolution. *Journal of Mathematical Biology* **65,** 553–580.

Crawford, F. W., Minin, V. N., and Suchard, M. A. (2014). Estimation for general birth-death processes. *Journal of the American Statistical Association* **109,** 730–747.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics* **25,** 631–650.

Daniels, H. E. (1982). The saddlepoint approximation for a general birth process. *Journal of Applied Probability* **19,** 20–28.

Darwin, J. H. (1956). The behaviour of an estimator for a simple birth and death process. *Biometrika* **43,** 23–31.

Davison, A. C. (2003). *Statistical Models*. Cambridge University Press, Cambridge.

Dion, J. P. and Yanev, N. M. (1994). Statistical inference for branching processes with an increasing random number of ancestors. *Journal of Statistical Planning and Inference* **39,** 329–351.

Duby, C. and Rouault, A. (1982). Estimation non paramétrique de l'espérance et de la variance de la loi de reproduction d'un processus de ramification. *Annales de l'Institut Henri Poincaré, XVIII Section B* **18,** 149–163.

Elliston, E. P. (1994). The black robin: Saving the world's most endangered bird. *BioScience* **44,** 499–501.

Guttorp, P. (1991). *Statistical Inference for Branching Processes*. Wiley, New York.

Harris, T. E. (2002). *The Theory of Branching Processes*. Dover, Mineola, NY.

Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes*, Volume 1. Academic Press, New York.

Keiding, N. (1975). Maximum likelihood estimation in the birth-and-death process. *Annals of Statistics* **3,** 363–372.

Kennedy, E. S. (2009). *Extinction vulnerability in two small, chronically inbred populations of Chatham Island black robin petroica traversi*. PhD thesis, Lincoln University, New Zealand.

Kvitkovičová, A. and Panaretos, V. M. (2011). Asymptotic inference for partially observed

branching processes. *Advances in Applied Probability*, **43,** 1166–1190.

Massaro, M., Sainudiin, R., Merton, D., Briskie, J. V., Poole, A. M., and Hale, M. L. (2013a). Human-assisted spread of a maladaptive behavior in a critically endangered bird. *PloS one* **8,** e79066.

Massaro, M., Stanbury, M., and Briskie, J. V. (2013b). Nest site selection by the endangered black robin increases vulnerability to predation by an invasive bird. *Animal Conservation* **16,** 404–411.

Massaro, M., Chick, A., Kennedy, E. S., and Whitsed, R. (2018). Post-reintroduction distribution and habitat preferences of a spatially limited island bird species. *Animal Conservation* **21,** 54–64.

Novozhilov, A. S., Karev, G. P., and Koonin, E. V. (2006). Biological applications of the theory of birth-and-death processes. *Briefings in Bioinformatics* **7,** 70–85.

Pedeli, X., Davison, A. C., and Fokianos, K. (2015). Likelihood estimation for the INAR(p) model by saddlepoint approximation. *Journal of the American Statistical Association* **110,** 1229–1238.

Ross, J. V., Taimre, T., and Pollett, P. K. (2006). On parameter estimation in population models. *Theoretical Population Biology* **70,** 498–510.

Xu, J., Guttorp, P., Kato-Maeda, M., and Minin, V. N. (2015). Likelihood-based inference for discretely observed birth-death-shift processes, with applications to evolution of mobile genetic elements. *Biometrics* **71,** 1009–1021.

Zhang, W., Bravington, M. V., and Fewster, R. M. (2019). Fast likelihood-based inference for latent count models using the saddlepoint approximation. *Biometrics* **75,** 723–733.

Supporting Information

Web Appendices A, B, C, D, and E, referenced in Sections 3, 4, 5, 7, and 8, and the MATLAB codes used to compute the saddlepoint approximations and parameter estimates via the different methods are available with this paper at the Biometrics website on Wiley Online Library. The data on the black robin population, the whooping crane population, and the H1N1 influenza outbreak, are available in separate MAT files alongside the other MATLAB files.

Received October 2007. *Revised February* 2008. *Accepted March* 2020.

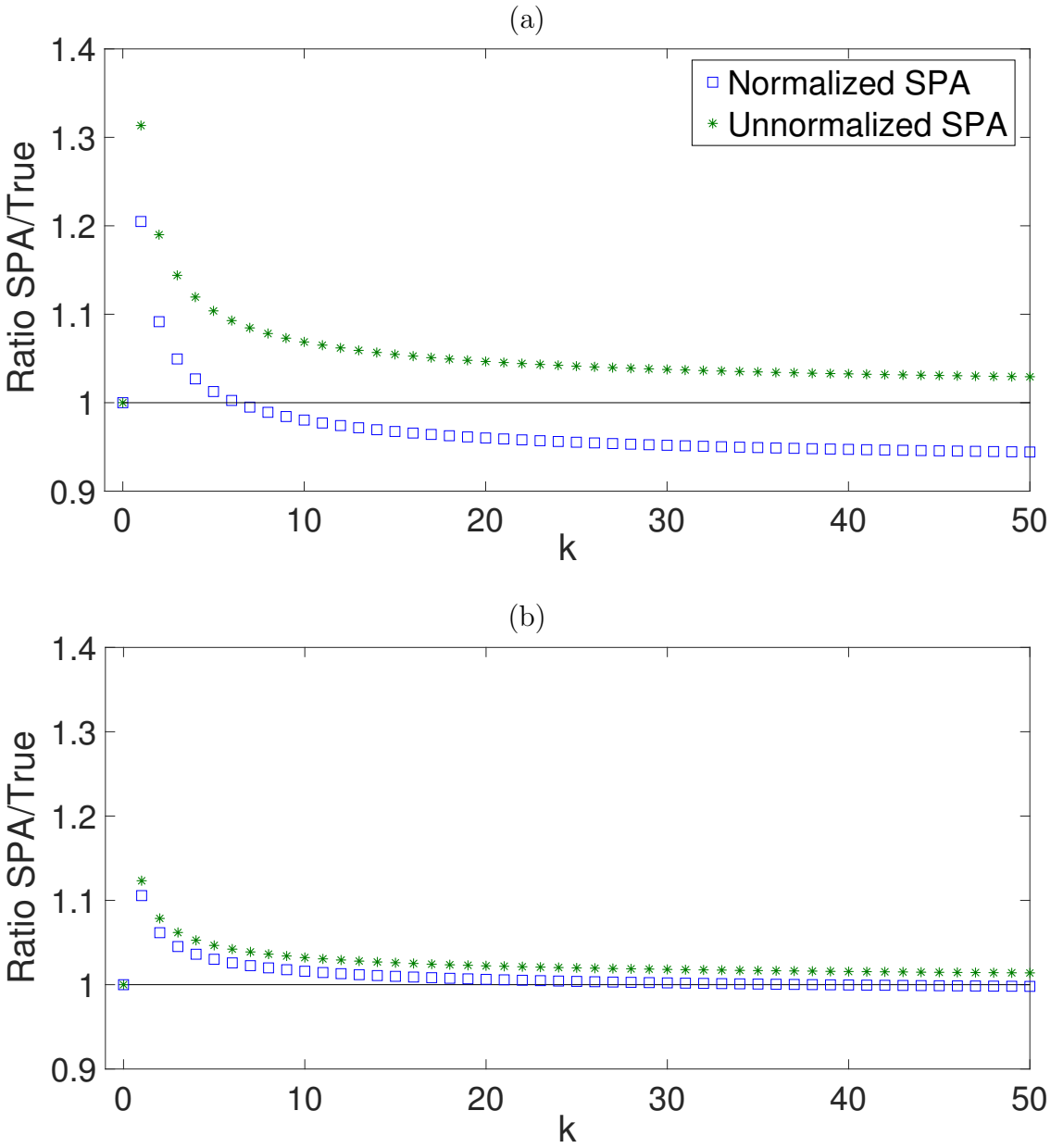


Figure 1. Ratio between the saddlepoint approximations and the exact probability mass function computed from (2) for the population size at $t = 1$ in an LBDP with $\lambda = 7$ and $\mu = 5$ and initial population sizes (a) $a = 5$, and (b) $a = 20$. Note that the $\square$ and $*$ symbols are superimposed at $k = 0$. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

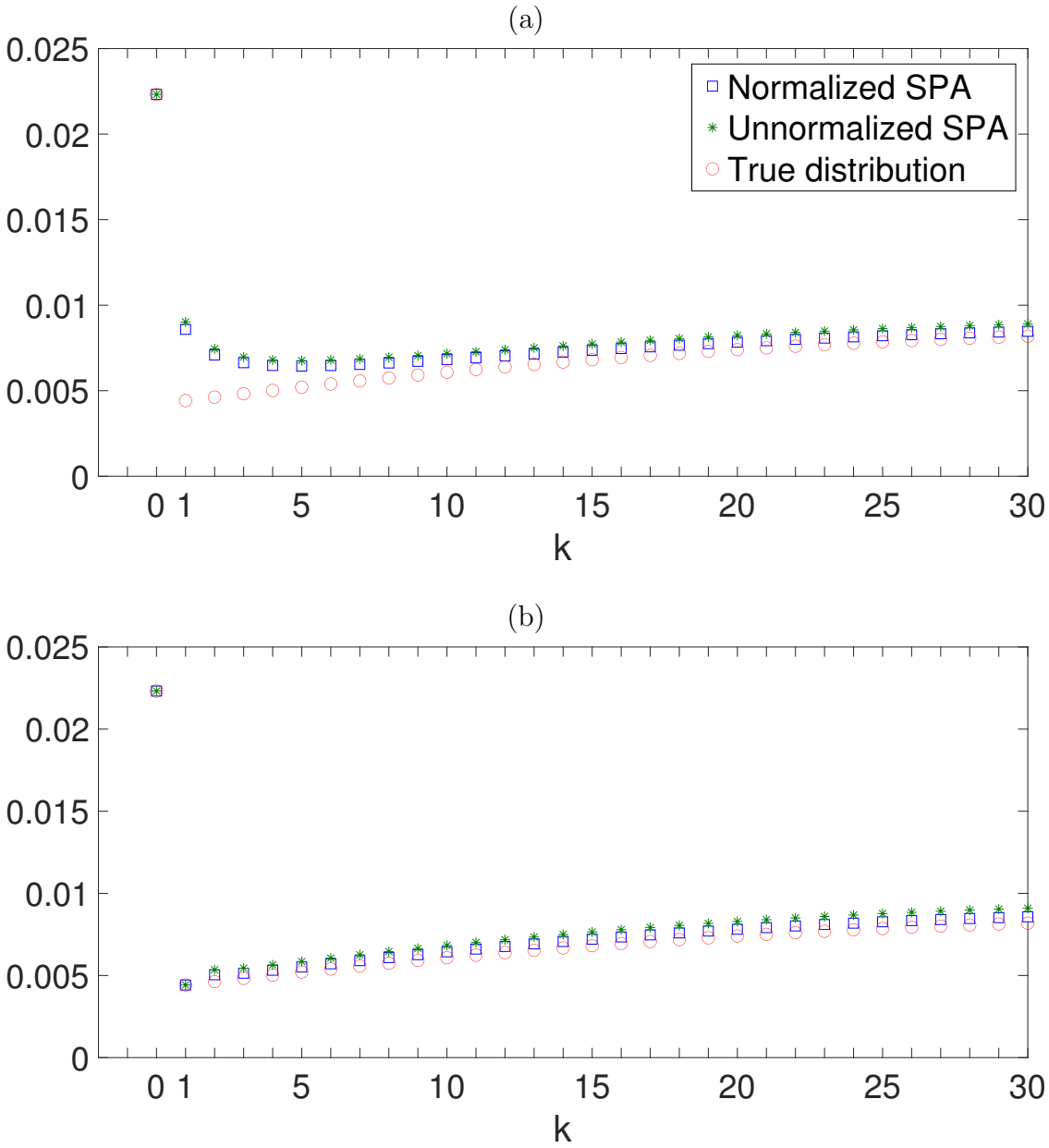


Figure 2. Probability mass function of the population size at $t = 1$ in an LBDP with $\lambda = 7$ and $\mu = 5$ and initial population size $a = 10$, together with its saddlepoint approximation $\tilde{p}_k(1; 10)$ and the normalized saddlepoint approximation $\bar{p}_k(1; 10)$. (a): unadjusted saddlepoint approximation; (b): adjusted saddlepoint approximation. Note that the $\square$, $*$, and $\circ$ symbols are superimposed at $k = 0$. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

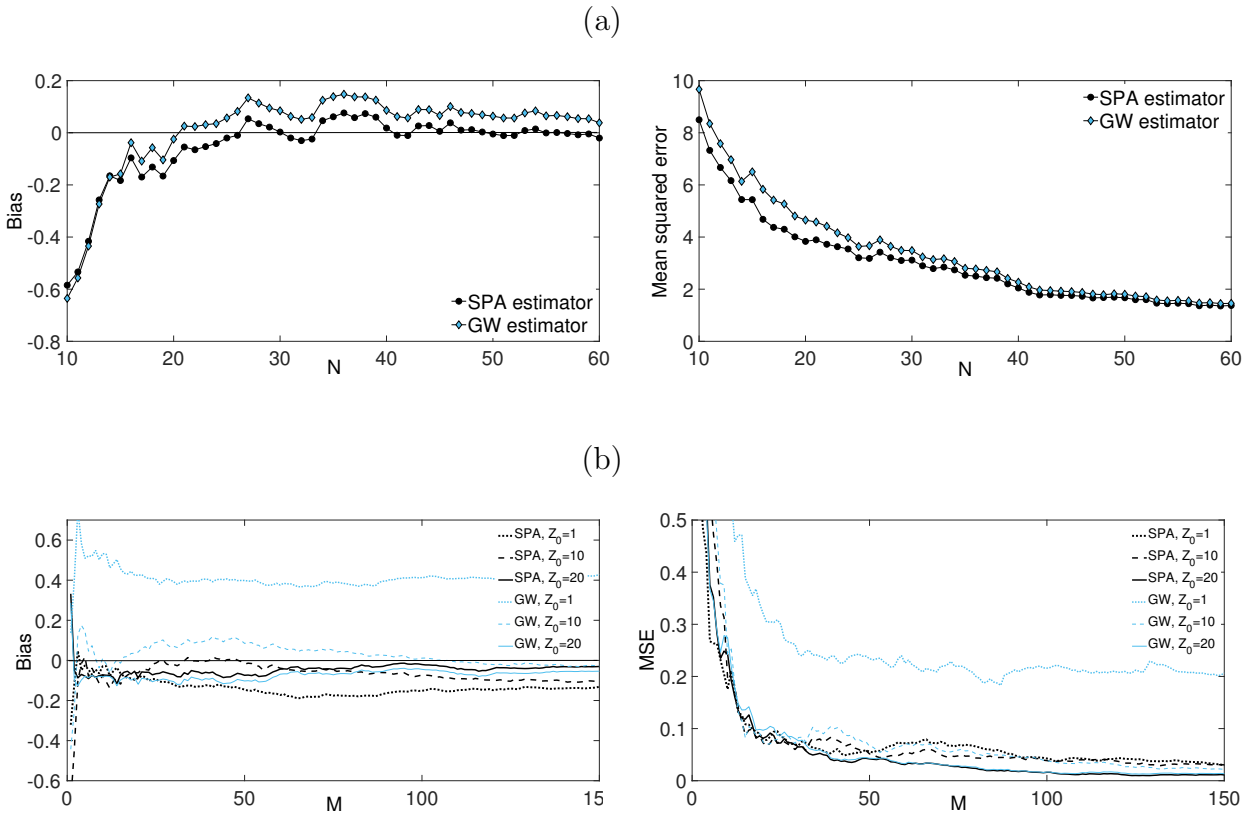


Figure 3. (a): Properties of $\widehat{\lambda}$ as N increases: Absolute bias, standard deviation and root mean square error for an LBDP with $\lambda = 7, \mu = 6$, $Z(0) = Z_0 = 10$. Results based on 100 simulations of a single ($M = 1$) non-extinct trajectory observed $N + 1$ times, at constant inter-observation intervals $\tau = 0.07$. (b): Properties of $\widehat{\lambda}$ as M increases: Absolute bias and root mean square error for an LBDP with $\lambda = 7, \mu = 5$, $Z(0) = Z_0 = 1, 10, 20$. Results based on 100 simulations of M non-extinct trajectories observed $N + 1 = 30$ times at constant inter-observation intervals $\tau = 1/10$. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

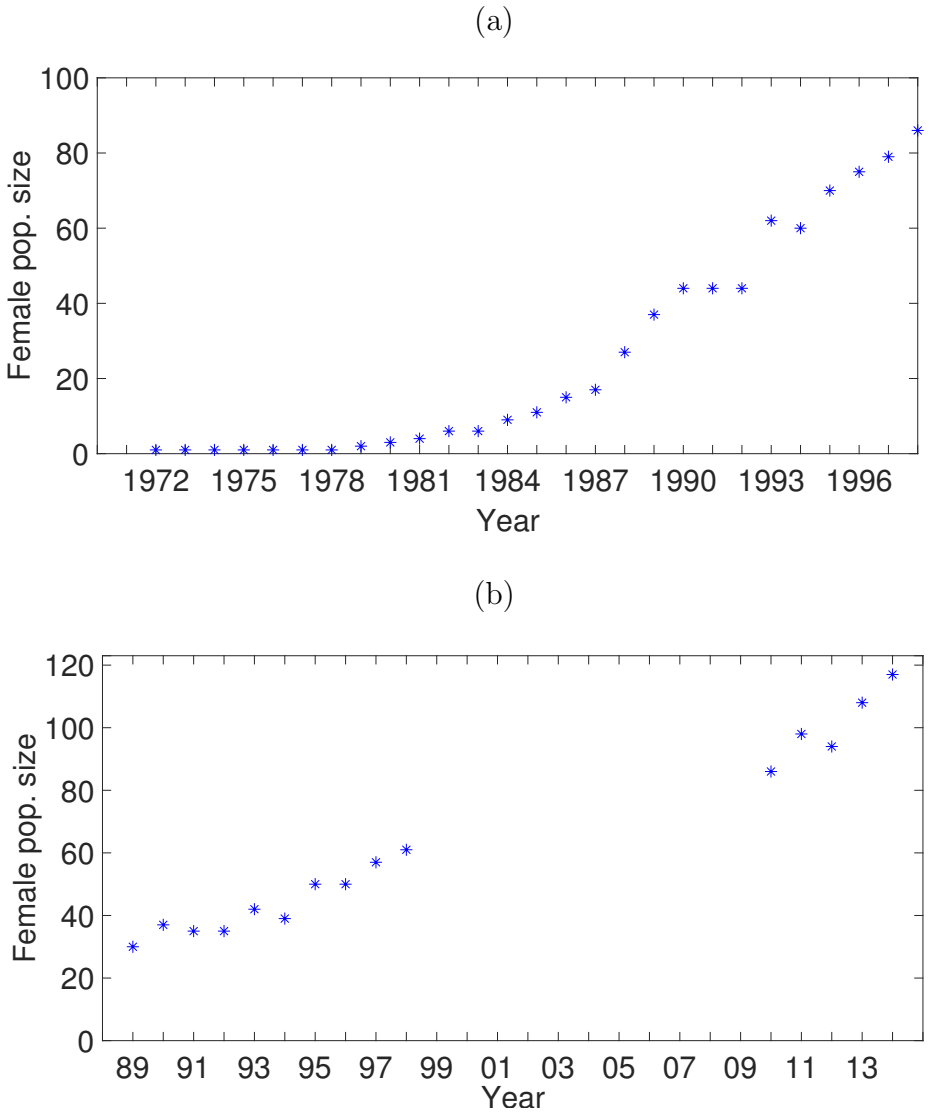


Figure 4. (a): Yearly black robin population census corresponding to females who survived up to at least one year, 1972–1998. (b): Female population on Rangatira Island, 1989–2014.

$M = 1$, $\mu = 6$, $\omega = 1$

Method	$Z_0 = 1$	$Z_0 = 5$	$Z_0 = 10$	$Z_0 = 20$	$Z_0 = 50$
SPMLE	2.40	2.44	2.40	2.45	2.48
GW	5.03	3.38	2.82	2.68	2.56

$M = 1$, $\mu = 5$, $\omega = 2$

Method	$Z_0 = 1$	$Z_0 = 5$	$Z_0 = 10$	$Z_0 = 20$
SPMLE	1.63	1.59	1.57	1.71
GW	2.15	1.75	1.64	1.62

$M = 20$, $\mu = 5$, $\omega = 2$

Method	$Z_0 = 1$	$Z_0 = 5$	$Z_0 = 10$	$Z_0 = 20$
SPMLE	0.39	0.35	0.35	0.36
GW	0.62	0.40	0.36	0.36

Table 1

Root mean square error of $\widehat{\lambda}$ (true value $\lambda = 7$) based on 5000 simulations of M non-extinct trajectories observed $N + 1 = 30$ times at constant inter-observation intervals $\tau = 1/10$ for different values of Z_0. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

(a)

	GW estimates	SPMLE	Adjusted SPMLE	MLE
$\widehat{\lambda}$	0.319 (0.071)	0.301 (0.073)	0.319 (0.072)	0.318 (0.072)
$\widehat{\mu}$	0.191 (0.071)	0.173 (0.072)	0.191 (0.071)	0.190 (0.071)
$\widehat{\omega}$	0.128 (0.028)	0.128 (0.027)	0.128 (0.028)	0.128 (0.028)
CPU est.	$2.67 \cdot 10^{-4}$	0.043	1.17	67.39
CPU SE	$4.61 \cdot 10^{-3}$	0.086	11.17	378.63

(b)

	Gen. GW	SPMLE	Adj. SPMLE	MLE
$\widehat{\lambda}$	0.2918 (0.0978)	0.2869 (0.0970)	0.2861 (0.0825)	0.2845 (0.0957)
$\widehat{\mu}$	0.2434 (0.1003)	0.2374 (0.0969)	0.2366 (0.0828)	0.2350 (0.0956)
$\widehat{\omega}$	0.0484 (0.0188)	0.0495 (0.0173)	0.0494 (0.0173)	0.0495 (0.0172)
CPU est.	0.0691	0.0564	0.6630	99.5229
CPU SE	1.4365	0.0731	4.9914	533.3362

Table 2

(a): Estimates (and standard errors based on Theorem 1 for the GW estimates, and on Fisher Information for the SPMLE and the MLE) of the birth, death, and growth rates of the black robin population between 1972 and 1998 from the GW, SPMLE and (exact) MLE approaches, and average CPU times (in seconds) to compute the estimates and the standard errors. (b): Same for the black robin population on Rangatira between 1989 and 2014 (with an eleven-year gap), where the GW estimates are replaced by the generalized GW estimates.