Consistent least squares estimation in population-size-dependent branching processes

Peter Braunsteins∗†, Sophie Hautphenne∗† Carmen Minuesa∗§

Abstract

We consider discrete-time parametric population-size-dependent branching processes (PSDBPs) with almost sure extinction and propose a new class of weighted least-squares estimators based on a single trajectory of population size counts. We prove that, conditional on non-extinction up to a finite time \( n \), our estimators are consistent and asymptotic normal as \( n \to \infty \). We pay particular attention to estimating the carrying capacity of a population. Our estimators are the first conditionally consistent estimators for PSDBPs, and more generally, for Markov models for populations with a carrying capacity. Through simulated examples, we demonstrate that our estimators outperform other least squares estimators for PSDBPs in a variety of settings. Finally, we apply our methods to estimate the carrying capacity of the endangered Chatham Island black robin population.

Keywords: branching process; population-size-dependence; almost sure extinction; inference; carrying capacity; least-squares estimation; consistency.

MSC 2020: 60J80.

1 Introduction and main results

Biological populations generally display logistic growth: after an initial period of exponential growth, competition for limited resources such as food, habitat, and breeding opportunities causes population growth to slow down until the population eventually reaches a carrying capacity, which is the maximum population size that a habitat can support. Biologists often aim to estimate the carrying capacity of the habitat of the population they are studying. One example is the Chatham Island black robin population (New Zealand), which was saved from the brink of extinction in the early 1980s when it was considered the world’s most endangered

∗All authors contributed equally to this work.
†School of Mathematics and Statistics, UNSW Sydney, NSW, 2052. email: p.braunsteins@unsw.edu.au, ORCID: 0000-0003-1864-0703
‡School of Mathematics and Statistics, The University of Melbourne, 3010 Parkville, VIC, Australia. email: sophiemb@unimelb.edu.au, ORCID: 0000-0002-8361-1901
§Departamento de Matemáticas, Instituto de Computación Científica Avanzada de la Universidad de Extremadura (ICCAEx), Facultad de Ciencias, Universidad de Extremadura (ROR: https://ror.org/0174shg90), 06006 Badajoz, Spain. email: cminuesaa@unex.es, ORCID: 0000-0002-8858-3145.
bird species [4]. The population has since recovered from a single breeding female to over 120 females, and appears to be nearing carrying capacity of its habitat (Rangatira Island) [18, 17]. In order to estimate the carrying capacity, ecologists often rely on assessing environmental characteristics of the habitat itself. For example, in [17] the carrying capacity of the black robin population is estimated to be 170 nesting pairs based on a habitat suitability model which requires precise knowledge of the habitat and is very costly. Here we take a different statistical approach to estimate the carrying capacity based solely on annual population size counts using stochastic population models.

In this paper we model biological populations with discrete-time population-size-dependent branching processes (PSDBPs), which can capture fundamental properties of many populations (including logistic growth). PSDBPs are Markov chains \( \{Z_n\}_{n \geq 0} \) characterised by the recursion

\[
Z_{n+1} = \sum_{i=1}^{Z_n} \xi_{n,i}(Z_n), \quad n \in \mathbb{N}_0 := \{0, 1, 2, \ldots\},
\]

where \( Z_n \) represents the population size at time \( n \) and \( \{\xi_{n,i}(\cdot)\}_{n,i} \) are independent random variables that represent each individual’s offspring whose distribution depends only on the current population size \( Z_n \). In these processes, state 0 is absorbing, and represents extinction of the population. Let \( m(z) := \mathbb{E}[\xi(z)] \) denote the mean offspring at population size \( z \); we say that a PSDBP has a carrying capacity \( K \) if

\[
m(z) > 1 \quad \text{when } z < K \quad \text{and} \quad m(z) < 1 \quad \text{when } z > K.
\]

When such a threshold value \( K \) exists, the PSDBP becomes extinct with probability one; this is common to almost all stochastic population models with a carrying capacity [11]. We point out that, while closely related, the mathematical definition of the carrying capacity in (2) is not the same as the biological definition given above.

The fact that PSDBPs with a carrying capacity become extinct with probability one introduces several technical challenges. For example, to perform an asymptotic analysis of an estimator \( \hat{\theta}_n \) based on a single trajectory, one needs to condition on non-extinction up to a finite time \( n \) — an event with vanishing probability as \( n \to \infty \). Consequently, an estimator \( \hat{\theta}_n \) may not satisfy the classical consistency property called \( C \)-consistency, defined as follows: an estimator \( \hat{\theta}_n := \hat{\theta}_n(Z_0, \ldots, Z_n) \) is called \( C \)-consistent for a parameter \( \theta \) if,

\[
\text{for any } \varepsilon > 0, \quad \lim_{n \to \infty} \mathbb{P}(|\hat{\theta}_n - \theta| > \varepsilon \mid Z_n > 0) = 0
\]

(see for instance [19]). In [2, Theorem 1] the authors show that the maximum likelihood estimators (MLEs) \( \hat{m}_n(z) \) (defined in [14]) for \( m(z) \) in a PSDBP are not \( C \)-consistent but instead are such that,

\[
\text{for any } \varepsilon > 0, \quad \lim_{n \to \infty} \mathbb{P}(|\hat{m}_n(z) - m^\dagger(z)| > \varepsilon \mid Z_n > 0) = 0,
\]

where \( m^\dagger(z) \) is different from \( m(z) \), and can be interpreted as the mean number of offspring born to individuals when the current population size is \( z \) in the \( Q \)-process associated with \( \{Z_n\} \), which corresponds to \( \{Z_n\} \) conditioned on non-extinction in the distant future (see Section 2.2). In [2] the authors refer to [14] as \( Q \)-consistency. A major gap remains to be filled in the context of PSDBPs with a carrying capacity: there does not exist any \( C \)-consistent estimator for \( m(z) \).
or for parameters (such as the carrying capacity) in parametric models. We point out that $C$-consistent estimators are lacking not only for PSDBPs but for any stochastic population process with a carrying capacity and an absorbing state at 0 (for example, the diffusion models in [16], the continuous-time birth-death processes in [9], and the controlled branching processes in [6]).

In this paper, we derive the first $C$-consistent estimators $\hat{\theta}_n$ for any set of parameters $\theta$ of parametric PSDBPs (under some regularity conditions); in particular, this leads to the first $C$-consistent estimators for the carrying capacity $K$. The key idea behind our $C$-consistent estimators is to embed the $Q$-consistent MLEs $\hat{m}_n(z)$ and their conditional limit $m^\uparrow(z, \theta)$ in the objective function of a weighted least squares estimator,

$$\tilde{\theta}_n := \arg\min_{\theta \in \Theta} \sum_{z \geq 1} \hat{w}_n(z) \left\{ \hat{m}_n(z) - m^\uparrow(z, \theta) \right\}^2,$$

with weights $\hat{w}_n(z) := \hat{w}_n(z; Z_0, \ldots, Z_n)$. As detailed in Section 3, this estimator differs from classical least squares estimators for stochastic processes (see for instance [13]) where the sum is taken over the successive times $i = 1, \ldots, n$ rather than the population sizes $z \geq 1$. To establish $C$-consistency and asymptotic normality of $\tilde{\theta}_n$, we build on the asymptotic properties of $\hat{m}_n(z)$ and overcome several new technical challenges which arise due to the almost sure extinction of the process, as well as the implicit form of the estimator $\tilde{\theta}_n$ and the fact that it includes an infinite sum.

Many populations, such as the Chatham Island black robins, are studied because they are still alive and conservation measures may need to be taken to preserve them. In these situations, the data should be viewed as being generated under the condition $Z_n > 0$. This suggests that for many populations (particularly small populations), $C$-consistent estimators may be desirable. We demonstrate the quality of our $C$-consistent estimators on simulated data in two contexts: (i) on growing populations that are yet to reach carrying capacity, and (ii) on (quasi-)stable populations that are fluctuating around the carrying capacity. In particular, we compare our $C$-consistent estimators $\tilde{\theta}_n$ with their counterpart obtained by replacing $m^\uparrow(z, \theta)$ with $m(z, \theta)$ in (5) (which is not $C$-consistent). We find that when the data include small population sizes (such as for the black robins), our $C$-consistent estimators $\tilde{\theta}_n$ outperforms its counterpart. Finally, we apply our methods to estimate the carrying capacity of the black robin population on Rangatira Island under several PSDBP models.

The paper is organised as follows. In the next section, we provide some background on PSDBPs and their associated $Q$-process, and we introduce a flexible class of parametric PSDBPs which will serve as an illustrative example. In Section 3 we introduce our weighted least squares estimators $\tilde{\theta}_n$ and in Section 4 we state their asymptotic properties. In Section 5, we compare our least squares estimators $\tilde{\theta}_n$ with their counterpart obtained by replacing $m^\uparrow(z, \theta)$ with $m(z, \theta)$ in (5) on simulated numerical examples. In Section 6 we apply our least squares estimators to the black robin data and estimate the carrying capacity of the population. Finally, Section 7 contains the proofs of our results.
2 Background

2.1 Parametric families of PSDBPs

A discrete-time population-size-dependent branching process (PSDBP) is a Markov chain \( \{Z_n\}_{n \in \mathbb{N}_0} \) characterised by the recursion given in \([1]\) with \( Z_0 = N \), where \( N \) is a positive integer, and where \( \{\xi_m(z) : i = 1, \ldots, z; \; z, n \in \mathbb{N}_0\} \) is a family of independent random variables with distribution \( p(z) = (p_k(z))_{k \in \mathbb{N}_0} \) such that \( p_k(z) = \mathbb{P}[^{\xi_m}_0(1) = k] \) depends on \( z \). We refer to \( p(z) \) as the offspring distribution at population size \( z \). In \( \{Z_n\} \) the state 0 is absorbing, and under minor regularity assumptions, all other states are transient \([10]\).

In this paper we assume that the family of offspring distributions \( (p(z))_{z \in \mathbb{N}} \) belongs to some parametric family, that is, we assume that there exists

\[
\theta_0 \in \text{int}(\Theta) \subseteq \mathbb{R}^d \quad \text{such that} \quad p(z) \equiv p(z, \theta_0), \quad \text{for each} \quad z \in \mathbb{N},
\]

where \( \text{int}(\Theta) \) denotes the interior of the parameter space \( \Theta \), and \( d < \infty \). As a consequence the PSDBP is parameterised by \( \theta_0 \). We assume that the map \( \theta \mapsto p(z, \theta) \) is a measurable function such that \( p_k(z, \cdot) \) is continuous on \( \Theta \) for each \( k \in \mathbb{N}_0 \) and \( z \in \mathbb{N} \), and is such that the offspring mean \( m(z) \equiv m(z, \theta) \) and variance \( \sigma^2(z) \equiv \sigma^2(z, \theta) \) are continuous on \( \Theta \) for each \( z \in \mathbb{N} \). In this framework, our aim is to estimate the parameter \( \theta_0 \) based on the sample of the population sizes \( Z_n := \{Z_0, \ldots, Z_n\} \).

2.2 The \( Q \)-process associated with \( \{Z_n\} \)

Let \( Q = (Q_{ij})_{i,j \in \mathbb{N}} \) refer to the sub-stochastic transition probability matrix of \( \{Z_n\}_{n \in \mathbb{N}_0} \) restricted to the transient states \( \{1, 2, \ldots\} \). The entry \( Q_{ij} \) is the probability that the population transitions from size \( i \) to size \( j \) in a single time unit. The ith row of \( Q \) thus corresponds to the i-fold convolution of the offspring distribution at population size \( i \). Because the family of offspring distributions \( (p(z))_{z \in \mathbb{N}} \) depends on \( \theta_0 \), so does the transition probability matrix, and consequently, \( Q \equiv Q(\theta_0) \), where this shorthand notation extends naturally to any \( \theta \in \Theta \).

We assume the following regularity conditions hold throughout the paper:

**Assumption 1.** For each \( \theta \in \Theta \),

- \((C1)\) The matrix \( Q(\theta) \) is irreducible.
- \((C2)\) \( \limsup_{z \to \infty} m(z, \theta) < 1 \).
- \((C3)\) For each \( v \in \mathbb{N} \), \( \sup_{z \in \mathbb{N}} \sum_{k=1}^{\infty} k^v p_k(z, \theta) < \infty \).

Under Assumptions \((C1),(C3)\) the PSDBP becomes extinct almost surely for any \( \theta \in \Theta \) and initial state \( Z_0 = i \) (see \([7\), Proposition 3.1\]), that is, \( \mathbb{P}^i(Z_n \to 0) = 1 \), where \( \mathbb{P}^i(\cdot) \) denotes the probability measure given the initial state is \( i \). In addition, there exists \( \rho, u, \) and \( v \) where \( \rho := \rho(Q) = \lim_{n \to \infty} (Q^n)_{ij}^{1/n} \), and \( u \) and \( v \) are strictly positive column vectors such that

\[
\begin{align*}
&u^\top Q = \rho u^\top, \quad Qv = \rho v, \quad u^\top 1 = 1, \quad \text{and} \quad u^\top v = 1, \quad (6) \\
\end{align*}
\]

and

\[
\begin{align*}
&Q^n \sim \rho^n v u^\top, \quad \text{as} \; n \to \infty; \quad (7)
\end{align*}
\]
see for instance [7]. The vector \( u \) corresponds to the quasi-stationary (or quasi-limiting) distribution of \( \{ Z_n \} \), and the vector \( v \) records the relative “strength” of each state. We refer the reader to [2, Section 2.3] for a more detailed explanation.

For every \( n \in \mathbb{N}_0 \) fixed, the process \( \{ Z_{\ell} \}_{0 \leq \ell \leq n} \) conditioned on \( Z_n > 0 \) is a time-inhomogeneous Markov chain that we denote by \( \{ Z_{\ell}^{(n)} \}_{0 \leq \ell \leq n} \) and whose one-step transition probabilities are

\[
\mathbb{P}(Z_{\ell+1}^{(n)} = j \mid Z_{\ell}^{(n)} = i) := \mathbb{P}(Z_{\ell+1} = j \mid Z_{\ell} = i, Z_n > 0) = Q_{ij} \frac{e_j^\top Q^{n-\ell-1} 1}{e_i^\top Q^{n-\ell} 1} \quad \text{for } i, j \geq 1, \quad (8)
\]

where \( e_i \) is a column vector whose \( i \)th entry is equal to 1 and whose remaining entries are 0. Taking the limit as \( n \to \infty \) in (8) and using (7) leads to homogeneous transition probabilities:

\[
\mathbb{P}(Z_{\ell+1} = j \mid Z_{\ell} = i) := \lim_{n \to \infty} \mathbb{P}(Z_{\ell+1}^{(n)} = j \mid Z_{\ell}^{(n)} = i) = \lim_{n \to \infty} Q_{ij} \frac{e_j^\top \rho^{n-\ell} v}{e_i^\top \rho^n v} = Q_{ij} \frac{v_j}{\rho v_i}. \quad (9)
\]

These transition probabilities define a new positive-recurrent time-homogeneous Markov chain \( \{ Z_{\ell}^\dagger \}_{\ell \geq 0} \), which we refer to as the \( Q \)-process associated with \( \{ Z_n \} \), and that can be interpreted as the original process \( \{ Z_n \} \) conditioned on not being extinct in the distant future. The \( n \)-step transition probabilities of the \( Q \)-process \( \{ Z_{\ell}^\dagger \} \) are given by

\[
(Q^n)_{ij} := \mathbb{P}(Z_{\ell}^\dagger = j \mid Z_0^\dagger = i) = (Q^n)_{ij} \frac{v_j}{\rho^i v_i}, \quad \text{for } n, i, j \geq 1,
\]

and, by (7), its stationary distribution is therefore

\[
\lim_{n \to \infty} \mathbb{P}(Z_{\ell}^\dagger = j \mid Z_0^\dagger = i) = u_j v_j, \quad \text{for } i, j \geq 1.
\]

For \( z \in \mathbb{N} \) and \( \theta \in \Theta \), we introduce the analogue of the mean offspring at population size \( z, m(z) \), in the \( Q \)-process:

\[
m^\dagger(z, \theta) = \frac{1}{z} \mathbb{E}[Z_{n+1, \theta}^\dagger \mid Z_n^\dagger = z] = \frac{1}{z} \sum_{k=1}^{\infty} k Q^\dagger_{zk}(\theta), \quad (10)
\]

and the analogue of the normalised offspring variance:

\[
\sigma^2(z, \theta) = \frac{1}{z^2} \text{Var}[Z_{n+1, \theta}^\dagger \mid Z_n^\dagger = z]
\]

\[
= \frac{1}{z^2} \left[ \sum_{k=1}^{\infty} k^2 Q^\dagger_{zk}(\theta) - \left( \sum_{k=1}^{\infty} k Q^\dagger_{zk}(\theta) \right)^2 \right] = \frac{1}{z^2} \sum_{k=1}^{\infty} k^2 Q^\dagger_{zk}(\theta) - m^\dagger(z, \theta)^2.
\]

### 2.3 A motivating example

In this section we introduce a flexible class of PSDBPs which we will return to in our empirical analysis in Section 5. We consider models in which, every time unit, each individual successfully reproduces with a probability \( r(z, \theta) \) that depends on the current population size \( z \), and if reproduction is successful, then the individual produces a random number of offspring with distribution \( b \) which does not depend on the current population size. More precisely, we consider
A parametric class of PSDBPs with offspring distributions \( \{p(z)\}_{z \in \mathbb{N}} \) which are characterised by

\[
p_j(z, \theta) = (1 - r(z, \theta)) \mathbf{1}_{\{j = 0\}} + r(z, \theta) b_j(\mu), \quad j \in \mathbb{N}_0,
\]

where \( z \in \mathbb{N}, b(\mu) = \{b_j(\mu)\}_{j \geq 0} \) is a distribution on the non-negative integers with mean \( \mu > 1 \), \( \theta = (K, \mu) \in \mathbb{R}_+ \times (1, \infty) \), and \( 0 \leq r(z, \theta) \leq 1 \). In other words, individuals give birth according to a zero-inflated distribution, where \( 1 - r(z, \theta) \) is the increase in the probability that individuals have no offspring. We can select the function \( r(\cdot) \) so that the PSDBP mimics well-established population models:

\[
r(z, \theta) = \frac{K}{K + (\mu - 1) z}, \quad \text{(Beverton-Holt model)}
\]

\[
r(z, \theta) = \left( \frac{1}{\mu} \right)^{z/K}, \quad \text{(Ricker model)}.
\]

For these choices of \( r(\cdot) \),

\[
m(z, \theta) > 1, \text{ if } z < K \quad \text{and} \quad m(z, \theta) < 1 \text{ if } z > K,
\]

and consequently \( K \) is the carrying capacity of the models. If the distribution \( b(\mu) \) is aperiodic and has finite moments, then Assumption 1 can be readily verified for the models described above.

To gain a better understanding of the Beverton-Holt and Ricker models we conduct a simulation study. We suppose the distribution \( b(\mu) \) is geometric. For each model, we fix the parameter \( \theta_0 = (K_0, \mu_0) = (40, 2) \), we suppose that the population starts with \( N = 1 \) individual, and we simulate \( N_0 = 250 \) independent non-extinct trajectories of the process for \( n = 5000 \) time units, that is, such that \( Z_n > 0 \). The first 150 time units of one of these trajectories for each model is plotted in Figure 1.

![Figure 1](image.png)

**Figure 1:** Evolution of the population sizes of two PSDBPs over the first 150 time units. Left: Beverton-Holt model. Right: Ricker model. Red solid line represents the carrying capacity.

Suppose that, based on the simulated trajectories, we would like to estimate the parameters \( \theta_0 \). To this end we recall the MLE for \( m(z, \theta) \), the mean number of offspring at population size \( z \), derived in [2, Proposition 1]:

\[
\hat{m}_n(z) := \frac{\sum_{i=0}^{n-1} Z_i + 1 \mathbf{1}_{\{Z_i = z\}}}{z \sum_{i=0}^{n-1} \mathbf{1}_{\{Z_i = z\}}},
\]

6
where \( \mathbb{1}_A \) denotes the indicator function of the set \( A \) and we take the convention that \( 0/0 = 0 \).

In Figure 2 we plot \( m(z, \theta_0), \hat{m}(z, \theta_0) \) (defined in (10)), and the mean of \( \hat{m}_n(z) \) for \( n = 5000 \) over the 250 trajectories, for the Beverton-Holt and the Ricker models. Figure 2 illustrates that the outcomes of \( \hat{m}_n(z) \) cling more tightly to the function \( m^\dagger(z, \theta_0) \) than to \( m(z, \theta) \), which is due to the bias caused by restricting our attention to non-extinct trajectories. In particular [2, Theorem 1] states that, for any initial population size \( i \in \mathbb{N} \) and population size \( z \in \mathbb{N} \),

\[
\forall \varepsilon > 0, \quad \lim_{n \to \infty} \mathbb{P}_i[|\hat{m}_n(z) - m^\dagger(z, \theta_0)| > \varepsilon | Z_n > 0] = 0 \quad (Q\text{-consistency}). \quad (15)
\]

In Figure 2 we see that the functions \( z \mapsto m^\dagger(z, \theta_0) \) (red curve) and \( z \mapsto m(z, \theta_0) \) (blue curve) are relatively close to each other at moderate and high population sizes \((z \geq 10)\), however, they are significantly different at low population sizes. This motivates the weighted least squares estimators that we introduce in the next section.

![Figure 2: Means of \( \hat{m}_n(z) \) (over the 250 processes) for each population size \( z \) observed. Left: Beverton-Holt model. Right: Ricker model. The red solid curve represents the function \( z \mapsto m^\dagger(z, \theta_0) \) while the blue solid curve represents the function \( z \mapsto m(z, \theta_0) \).](image)

### 3 Weighted least squares estimators

In the context of branching processes, classical least squares estimators minimise the (weighted) squared differences between the population sizes \( Z_k \) and their conditional mean \( Z_{k-1} m(Z_{k-1}, \theta) \), for \( k = 1, \ldots, n \):

\[
\hat{\theta}_n^* := \arg \min_{\theta \in \Theta} \sum_{k=1}^{n} w(Z_{k-1}) \{ Z_k - Z_{k-1} m(Z_{k-1}, \theta) \}^2,
\]

where \( w(\cdot) \) is an appropriately chosen weighting function (such as \( w(Z_{k-1}) = Z_{k-1}^{-1} \), see [8, p. 40]). This approach has been applied to Galton–Watson branching processes [8, Section 2.3], multi-type controlled branching processes [5], branching processes with immigration [13], and PSDBPs [14, 15]. Each of these papers develops asymptotic properties of \( \hat{\theta}_n^* \), such as consistency on the set of non-extinction, for processes with a positive chance of survival. Here we consider branching processes that become extinct almost surely, which introduces several technicalities that lead us to take a different approach.

To estimate \( \theta_0 \) we develop a novel class of least squares estimators that use the \( Q\text{-consistent} \) estimators \( \{ \hat{m}_n(z) \}_{z \in \mathbb{N}} \), and that are \( C\text{-consistent} \), that is, that satisfy [3]. The idea is that,
instead of summing over all time units \( k = 1, \ldots, n \) as in \((16)\), we sum over all population sizes \( z \geq 1 \) and minimise the (weighted) squared differences between the MLE \( \hat{m}_n(z) \) and its conditional limit \( m^\dagger(z, \theta) \):

\[
\hat{\theta}_n := \arg \min_{\theta \in \Theta} \sum_{z=1}^{\infty} \hat{w}_n(z) \left\{ \hat{m}_n(z) - m^\dagger(z, \theta) \right\}^2, 
\]

where the weights \( \{\hat{w}_n(z)\}_{z \in \mathbb{N}} \) are computed from the observations \( Z_0, Z_1, \ldots, Z_n \). Estimators in this class can be viewed as hybrids between an MLE and a least squares estimator, where the MLEs \( \{\hat{m}_n(z)\}_{z \in \mathbb{N}} \) first summarise the data before the least squares step is applied. In Figure 3 we give a visual representation of the classical least squares approach (top panel) and the hybrid approach taken in this paper (bottom panel).

![Figure 3: Top panel: the blue points correspond to a trajectory of \( \{Z_n\} \) simulated from a PSDBP with carrying capacity \( K = 30 \). The red curve represents the conditional expectation of \( Z_n \) given the value of \( Z_{n-1} \). This panel illustrates the differences \( Z_k - Z_{k-1} m(Z_{k-1}, \theta_0) = Z_k - E[Z_k | Z_{k-1}] \) in the classical least squares estimator \( \hat{\theta}_n \). Bottom panel: The trajectory of \( Z_n \) is summarised in the values of \( \{\hat{m}_n(z)\}_{z \geq 1} \) represented by the blue points. The green curve represents the function \( m^\dagger(z, \theta_0) \). This panel illustrates the differences \( \hat{m}_n(z) - m^\dagger(z, \theta) \) in the least squares estimator \( \hat{\theta}_n \).](image)

We suppose that the weights in \((5)\) satisfy the following assumption:
(A1) For each \( n \in \mathbb{N} \), \( \{\hat{w}_n(z)\} \) is an empirical distribution (i.e., for each \( z \in \mathbb{N} \), \( \hat{w}_n(z) \geq 0 \), and \( \sum_{z \geq 1} \hat{w}_n(z) = 1 \)), and there exists a limiting distribution \( \{w_z \equiv w_z(\theta_0)\}_{z \in \mathbb{N}} \) such that, for any \( i, z \geq 1 \) and \( \varepsilon > 0 \),
\[
\lim_{n \to \infty} P_i[|\hat{w}_n(z) - w_z| > \varepsilon | Z_n > 0] = 0.
\]

Two natural choices for the weight functions are
\[
\hat{w}_n^{(1)}(z) := \frac{\sum_{i=0}^{n-1} 1\{Z_i = z\}}{n} \quad \text{and} \quad \hat{w}_n^{(2)}(z) := \frac{z \sum_{i=0}^{n-1} 1\{Z_i = z\}}{\sum_{i=0}^{n-1} Z_i},
\]
where \( \hat{w}_n^{(1)}(z) \) assigns weights according to the proportion of time units with population size \( z \), and \( \hat{w}_n^{(2)}(z) \) assigns weights according to the proportion of parents alive when the population size is \( z \).

Lemma 1. The weights \( \{\hat{w}_n^{(1)}(z)\} \) and \( \{\hat{w}_n^{(2)}(z)\} \) satisfy Assumption (A1) with respective limiting distributions \( \{w_z^{(1)}\} \) and \( \{w_z^{(2)}\} \) characterised by
\[
w_z^{(1)} = u_z v_z \quad \text{and} \quad w_z^{(2)} = \frac{z u_z v_z}{\sum_{k=1}^{\infty} k u_k v_k}, \quad z \in \mathbb{N}.
\]

Note that \( \{w_z^{(1)}\} \) is the stationary distribution of the \( Q \)-process \( \{Z_\ell^i\} \), and \( \{w_z^{(2)}\} \) is the size-biased distribution associated with \( \{w_z^{(1)}\} \).

### 4 Asymptotic properties

We now establish that, under regularity assumptions, the class of least squares estimators \( \hat{\theta}_n \) are \( C \)-consistent and asymptotically normal. We start by introducing some notation: let \( \| \cdot \| \) denote any norm in \( \mathbb{R}^d \) (recall that \( \theta \) is a \( d \)-dimensional parameter), let \( \Psi_{\Sigma}(\cdot) \) denote the distribution function of a multivariate normal distribution with mean vector \( 0 = (0, \ldots, 0)^\top \) and covariance matrix \( \Sigma \) whose dimension will be clear from the context, and let \( \nabla \) denote the gradient operator with respect to \( \theta \).

We make the following assumptions:

(A2) If \( m^\dagger(z, \theta_1) = m^\dagger(z, \theta_2) \) for each \( z \in \text{supp}(w(\theta_0)) \), then \( \theta_1 = \theta_2 \).

(A3) \( M := \sup_{\theta \in \Theta} \sup_{z \in \mathbb{N}} m^\dagger(z, \theta) < \infty \).

(A4) For any initial state \( i \in \mathbb{N} \), there exists \( C > 0 \) such that
\[
\lim_{n \to \infty} P_i \left[ \sup_{z \in \mathbb{N}} |\hat{m}_n(z) - m^\dagger(z, \theta_0)| < C \mid Z_n > 0 \right] = 1.
\]

(A5) The function \( m^\dagger(z, \theta) \) is twice continuously differentiable with respect to \( \theta \) for each \( z \in \mathbb{N} \). Moreover, for each \( \theta' \in \Theta \), there exists a compact set \( \mathcal{C} \equiv \mathcal{C}(\theta') \) such that \( \theta' \in \text{int}(\mathcal{C}) \) and
(a) $M_1^*(\theta') := \sup_{\theta \in C} \sup_{z \in \mathbb{N}} \| \nabla m^\uparrow(z, \theta) \|_\infty < \infty.$

(b) $M_2^*(\theta') := \sup_{\theta \in C} \sup_{z \in \mathbb{N}} \max_{i,j = 1, \ldots, d} \left| \frac{\partial^2 m^\uparrow(z, \theta)}{\partial \theta_i \partial \theta_j} \right| < \infty.$

We comment on these assumptions in Remark 1 below.

**Theorem 1** (C-consistency of $\hat{\theta}_n$). Under Assumptions [(A1)-(A4)], for any $i \in \mathbb{N},$

(i) Existence:
\[
\lim_{n \to \infty} \mathbb{P}_i \left[ \exists \hat{\theta}_n \mid Z_n > 0 \right] = 1.
\]

(ii) C-consistency:
\[
\forall \varepsilon > 0, \quad \lim_{n \to \infty} \mathbb{P}_i \left[ \| \hat{\theta}_n - \theta_0 \| > \varepsilon \mid Z_n > 0 \right] = 0.
\]

**Theorem 2** (C-normality of $\hat{\theta}_n$). Assume that either

(i) there exists a value $z^* < \infty$ such that for all $n \in \mathbb{N}$ and $z \geq z^*, \hat{w}_n(z) = 0,$ and Assumption [(A1)] holds; or

(ii) for all $n \in \mathbb{N}, \{\hat{w}_n(z)\} = \{\hat{w}^{(1)}_n(z)\};$ or

(iii) for all $n \in \mathbb{N}, \{\hat{w}_n(z)\} = \{\hat{w}^{(2)}_n(z)\}.$

Then, under Assumptions [(A2)-(A5)], for any $i \in \mathbb{N}$ and $x_1, \ldots, x_d \in \mathbb{R},$
\[
\lim_{n \to \infty} \mathbb{P}_i \left[ \sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \in (-\infty, x_1] \times \ldots \times (-\infty, x_d] \mid Z_n > 0 \right] = \Psi_{\beta(\theta_0)}(x_1, \ldots, x_d),
\]
where
\[
\beta(\theta_0) = \eta(\theta_0)^{-1} \zeta(\theta_0) \eta(\theta_0)^{-1}
\]
is a positive semi-definite matrix, with the $d$-dimensional matrices $\eta(\theta_0)$ and $\zeta(\theta_0)$ given by
\[
\eta(\theta_0) = 2 \sum_{z=1}^{\infty} w_z(\theta_0) \nabla m^\uparrow(z, \theta_0) \nabla m^\uparrow(z, \theta_0)^\top
\]
\[
\zeta(\theta_0) = 4 \sum_{z=1}^{\infty} w_z(\theta_0)^2 \gamma(z, \theta_0) \nabla m^\uparrow(z, \theta_0) \nabla m^\uparrow(z, \theta_0)^\top,
\]
where
\[
\gamma(z, \theta_0) := \frac{\sigma^{2\uparrow}(z, \theta_0)}{u_z(\theta_0)v_z(\theta_0)}.
\]

**Remark 1.** Assumption (A2) is expected to hold for a wide a variety of models including those considered in Section 2.3. Assumptions (A3) and (A4) hold if, for instance, the offspring distribution has bounded support, which is a reasonable assumption in most biological models. Assumption (A5) contains technical conditions that, in practice, would likely need to be verified numerically.

**Remark 2.** Applying Theorem 2 with weights that satisfy (i) for a large value of $z^*$ should be sufficient in practice. Indeed, under Assumption [(C2)], the process rarely takes very large values.
5 Simulated examples

In this section we compare the $C$-consistent estimator $\hat{\theta}_n$ (defined in (5)) with its natural counterpart obtained by replacing $m^*(z, \theta)$ with $m(z, \theta)$ in (5):

$$\tilde{\theta}_n := \arg \min_{\theta \in \Theta} \sum_{z=1}^{\infty} \hat{w}_n(z) \left\{ \hat{m}_n(z) - m(z, \theta) \right\}^2.$$

(20)

In general, due to (4), $\tilde{\theta}_n$ is not $C$-consistent when extinction is almost sure. In our examples, we consider the two weighting functions $\{\hat{w}_n^{(1)}(z)\}$ and $\{\hat{w}_n^{(2)}(z)\}$ defined in (17).

We separate our analysis into two cases: (i) populations which are still growing and are yet to reach the carrying capacity, and (ii) populations that start in the quasi-stationary regime (near the carrying capacity). We compare the accuracy of $\hat{\theta}_n$ and $\tilde{\theta}_n$ based on non-extinct trajectories.

5.1 Growing populations

Consider the illustrative example in Section 2.3 whose offspring distribution is given by (11), with $r(z, \theta)$ given by (12) (Beverton-Holt model) and where $b(\mu)$ is a binary splitting distribution with $b_2 = v$ and $b_0 = 1 - v$ (so that $\mu = 2v$). We fix the parameters $\theta_0 = (K_0, v_0) = (100, 0.6)$ and the initial population size $N_0 = 2$. We simulate $N_0 = 2000$ trajectories of the process for $n = 15, 20, 25$ time units such that $Z_n > 0$. We estimate the parameters $\theta_0$ based on the observed population sizes $z_0, \ldots, z_n$. We are interested in comparing the performance of the estimators in the challenging —but practically important (see Section 6)— setting where the largest observation in most of the trajectories is well below the carrying capacity.

In Figure 4 we illustrate the sample mean of the 2000 trajectories (left) and the histogram of the population size $Z_{25}$ of the simulated paths (right). We remark that the sample mean of $Z_{25}$ is well below the carrying capacity $K_0 = 100$ and its variance is large.

![Image](image-url)

Figure 4: Beverton-Holt model with $(K_0, v_0) = (100, 0.6)$. Left: Mean path of the 2000 simulated trajectories. Right: Histogram of $Z_{25}$; the red vertical line is the sample mean and the black vertical line is $K_0 = 100$. 
In Table 1, we provide the mean, median, standard deviation (SD) and relative mean squares error (RMSE) of the estimates for $K_0$ and $v_0$ using the C-consistent estimator $\hat{\theta}_n = (\hat{K}_n, \hat{v}_n)$ in (5) and its counterpart $\tilde{\theta}_n = (\tilde{K}_n, \tilde{v}_n)$ in (20) with weighting functions $\{\hat{w}_n^{(1)}(z)\}$ and $\{\hat{w}_n^{(2)}(z)\}$ defined in (17), for $n = 10, 15, 20$.

Table 1: Beverton-Holt model with $(K_0, v_0) = (100, 0.6)$. Mean, median, standard deviation (SD) and relative mean squares error (RMSE) of the estimates for $K_0$ and $v_0$ using the C-consistent estimator $\theta_n = (K_n, v_n)$ in (5) and its counterpart $\tilde{\theta}_n = (\tilde{K}_n, \tilde{v}_n)$ in (20) with weighting functions $\{\hat{w}_n^{(1)}(z)\}$ and $\{\hat{w}_n^{(2)}(z)\}$ defined in (17), for $n = 10, 15, 20$. First, we generally observe that, as expected, the estimates improve as $n$ increases. Second, we observe that the weighting function $\{\hat{w}_n^{(2)}(z)\}$ generally outperforms $\{\hat{w}_n^{(1)}(z)\}$; this is because the weights should ideally be equal to the inverse of the variance of the errors, and the limiting expression $w_z^{(2)}$ in Lemma 1 is (approximately) inversely proportional to $\gamma(z, \theta_0)$ in (19) which can be interpreted as the variance of the error $\{\hat{m}_n(z) - m(z, \theta)\}$ in the $z$th term of the least squares estimator $\hat{\theta}_n$ (see (32)).

More interestingly, for all $n$ and for both $K$ and $v$, the mean and median of $\tilde{\theta}_n$ are closer to the true value $\theta_0 = (100, 0.6)$ than the counterpart $\hat{\theta}_n$ for any fixed weighting function; this illustrates that $\tilde{\theta}_n$ has a smaller bias than $\hat{\theta}_n$. The relative mean squares error is also always lower for $\tilde{\theta}_n$ than for $\hat{\theta}_n$. On the other hand, the standard deviation and relative mean squares error are smaller for $\tilde{K}_n$ than for $\hat{K}_n$. This is due to the presence of more outliers among the values of $\tilde{K}_n$ than among those of $\hat{K}_n$. In the left panel of Figure 5, we display the values of the difference $\hat{K}_n - \tilde{K}_n$; we see that these differences are generally positive, which decreases the overall bias of $\hat{K}_n$ (recall that the median of $\hat{K}_{25}(w_3^{(2)})$ is 80.255), but there is a small chance that the difference is very large, which causes the standard deviation and relative mean squares error of $\hat{K}_n$ to be larger than those of $\tilde{K}_n$.

The box plots of $\hat{\theta}_n$ and $\tilde{\theta}_n$ with the weighting function $\{\hat{w}_n^{(2)}(z)\}$ and $\{\hat{w}_n^{(1)}(z)\}$ are illustrated in Figures 6 and 7 respectively. In these figures, we see that, for both $\{\hat{w}_n^{(2)}(z)\}$ and $\{\hat{w}_n^{(1)}(z)\}$, the true value of $K$ and $v$ generally lies between the first and third quantiles of the empirical distribution of $\hat{\theta}_n$, while this is not generally the case for $\tilde{\theta}_n$. In addition, we see that there is a bigger difference between $\hat{\theta}_n$ and $\tilde{\theta}_n$ for the weighting function $\{\hat{w}_n^{(1)}(z)\}$ than for $\{\hat{w}_n^{(2)}(z)\}$. This is because there are larger differences between $m(z)$ and $m(\hat{z})$ for small values of $z$ (recall Figure 2), and $\{\hat{w}_n^{(1)}(z)\}$ places more weight on small values of $z$ than $\{\hat{w}_n^{(2)}(z)\}$.

Finally, in Table 2, we present the estimates obtained when the 2000 trajectories are accumulated into a single sample, that is, we take

$$\hat{m}_n(z) = \frac{\sum_{j=1}^{2000} \sum_{i=0}^{n-1} Z_{j,i+1} I\{Z_{j,i} = z\}}{z \sum_{j=1}^{2000} \sum_{i=0}^{n-1} I\{Z_{j,i} = z\}}.$$
Figure 5: Beverton-Holt model with \((K_0, v_0) = (100, 0.6)\). Marginal empirical distributions of the difference in the outcomes of the estimators \(\hat{\theta}_n\) and \(\tilde{\theta}_n\), with weighting function \(\{\tilde{w}_n^{(2)}(z)\}\) for the 2000 simulated trajectories, at \(n = 25\).

| Method                    | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|---------------------------|----|----|----|----|----|----|----|----|----|----|----|
| \(\hat{K}_n(w^{(1)})\)  | 87.760 | 90.122 | 92.824 | 94.639 | 96.776 | 97.665 | 98.167 | 97.805 | 98.457 | 98.040 | 98.067 |
| \(\hat{K}_n(w^{(1)})\)  | 45.730 | 49.477 | 53.353 | 56.905 | 60.421 | 63.525 | 66.340 | 68.613 | 71.073 | 72.981 | 74.810 |
| \(\hat{K}_n(w^{(2)})\)  | 91.781 | 94.270 | 97.384 | 99.505 | 100.856 | 101.601 | 101.296 | 100.737 | 101.186 | 100.437 | 100.302 |
| \(\hat{K}_n(w^{(2)})\)  | 72.118 | 76.681 | 81.345 | 85.115 | 88.306 | 90.636 | 92.039 | 92.818 | 94.206 | 94.502 | 95.104 |
| \(\hat{v}_n(w^{(1)})\)  | 0.607 | 0.606 | 0.606 | 0.605 | 0.604 | 0.604 | 0.604 | 0.604 | 0.604 | 0.604 | 0.604 |
| \(\hat{v}_n(w^{(2)})\)  | 0.687 | 0.682 | 0.674 | 0.671 | 0.668 | 0.666 | 0.664 | 0.662 | 0.660 | 0.659 | 0.659 |
| \(\hat{v}_n(w^{(2)})\)  | 0.604 | 0.604 | 0.603 | 0.602 | 0.601 | 0.601 | 0.601 | 0.601 | 0.601 | 0.601 | 0.601 |
| \(\hat{v}_n(w^{(2)})\)  | 0.633 | 0.629 | 0.626 | 0.623 | 0.621 | 0.619 | 0.618 | 0.617 | 0.616 | 0.615 | 0.614 |

Table 2: Beverton-Holt model with \((K_0, v_0) = (100, 0.6)\). Estimates obtained when the 2000 trajectories are accumulated into a single sample.

where \(Z_{j,t}\) is the population size at time \(t\) in the \(j\)th trajectory, and we adjust the weighting functions \(\{\tilde{w}_n^{(2)}(z)\}\) and \(\{\tilde{w}_n^{(1)}(z)\}\) similarly. We observe that \(\hat{\theta}_n\) significantly outperforms \(\tilde{\theta}_n\) with both weighting functions. While the \(C\)-consistency of \(\hat{\theta}_n\) stated in Theorem 1 ensures that \(m'(z, \theta)\) is the correct value to subtract in our least squares estimator for large values of \(n\), Table 2 strongly suggests that this is also the case for relatively small values of \(n\).

5.2 Quasi-stationary populations

We consider the same Beverton-Holt model as in Section 5.1 and now focus on populations that linger around the carrying capacity. We fix \(K_0 = 50\) and consider two scenarios, one where there are small fluctuations of \(Z_n\) around the carrying capacity \((v_0 = 0.7)\), and one where there are larger fluctuations \((v_0 = 0.54)\). The amplitude of the fluctuations depends on the parameter \(\mu = 2v\) as follows: the larger the absolute value of the derivative of the mean offspring function \(m(z)\) (assuming its support is extended to \(\mathbb{R}_+\)) at \(z = K\), the smaller the fluctuations, as this derivative controls the strength of the drift toward \(K\). For the Beverton-Holt model, \(m'(z)|_{z=K} = 1 - 2v\), so taking \(v_0 = 0.7\) leads to smaller fluctuations than \(v_0 = 0.54\).

We fix the initial population size \(N = 2\) and simulate \(N_0 = 2000\) trajectories of the process for \(n = 50, 500, 2000\) time units such that \(Z_n > 0\). In Figure 8 we plot the histogram of the population size \(Z_{2000}\) based on the simulated trajectories for \(v_0 = 0.7\) (left panel) and \(v_0 = 0.54\).

13
Figure 6: Beverton-Holt model with \((K_0, v_0) = (100, 0.6)\). Box plots of \(\hat{\theta}_n\) and \(\tilde{\theta}_n\) with the weighting function \(\{\hat{w}_n^{(2)}(z)\}\). Red lines represent the true value of each parameter.

Figure 7: Beverton-Holt model with \((K_0, v_0) = (100, 0.6)\). Box plots of \(\hat{\theta}_n\) and \(\tilde{\theta}_n\) with the weighting function \(\{\hat{w}_n^{(1)}(z)\}\). Red lines represent the true value of each parameter.
Figure 8: Beverton-Holt model. Left: Case with \((K_0, v_0) = (50, 0.7)\). Right: Case with \((K_0, v_0) = (50, 0.54)\). Red solid line is the mean population size at time \(n = 2000\) over the 2000 simulated paths. Black solid line is the carrying capacity \(K_0\).

(right panel). The spread of the two histograms confirms that the population exhibits larger fluctuations around the carrying capacity when \(v_0 = 0.54\) than when \(v_0 = 0.7\).

We plot the histograms of the estimates for \(K_0\) and \(v_0\) and the confidence regions of their joint empirical distribution when \(n = 2000\) and \(K_0 = 50\), for \(v_0 = 0.7\) in Figure 9 and for \(v_0 = 0.54\) in Figure 10. The blue curve plotted in the histograms for \(\hat{K}\) and \(\hat{v}\) is the density of the normal distribution with mean and variance computed using Theorem 2. The vertical blue lines give the boundaries of the resulting theoretical 95\% confidence intervals, and the vertical dashed lines give the boundaries of the empirical 95\% confidence intervals. In Tables 3 and 4 we provide the mean, median, standard deviation (SD) and relative mean square error (RMSE) of the estimates obtained with each of the estimators.

From both the figures and the tables, we observe that when \(v_0 = 0.7\) (small fluctuations around the carrying capacity), the distributions of \(\hat{\theta}_n\) and \(\hat{\theta}_n\) are similar across both weighting functions. In contrast, when \(v_0 = 0.54\) (larger fluctuations around the carrying capacity), there is a visible difference between the distributions of \(\hat{\theta}_n\) and \(\tilde{\theta}_n\), and there is a clear bias in \(\tilde{\theta}_n\). To understand why, recall that there is a larger difference between \(m^\dagger(z, \theta)\) and \(m(z, \theta)\) for small values of \(z\), and when \(v_0 = 0.54\) there are larger fluctuations around the carrying capacity, which means that the process reaches these lower population sizes more frequently.

The simulated examples in this section help us to form a general conclusion: if we are
Figure 9: Case \((K_0, v_0) = (50, 0.7)\). Left: confidence regions for levels 50%, 75%, 90%, 95%, 97.5%. Centre: marginal distribution of the estimator of \(K\), together with the empirical and theoretical marginal 95% confidence intervals. Right: marginal distribution of the estimator of \(v\), together with the empirical and theoretical marginal 95% confidence intervals.
Figure 10: Case $(K_0, v_0) = (50, 0.54)$. Left: confidence regions for levels 50%, 75%, 90%, 95%, 97.5%. Centre: marginal distribution of the estimator of $K$, together with the empirical and theoretical marginal 95% confidence intervals. Right: marginal distribution of the estimator of $v$, together with the empirical and theoretical marginal 95% confidence intervals.
there are no observations at small population sizes, then a Beverton-Holt model with

\[ \theta_n \]

Table 4: Beverton-Holt model with \((K_0, v_0) = (50, 0.54)\). Mean, median, standard deviation (SD) and relative
mean square error (RMSE) of the estimates obtained with each of the estimators.

| \(n = 50\)     | \(\hat{K}_n(w^{(1)}_n)\) | \(\hat{K}_n(w^{(2)}_n)\) | \(\tilde{K}_n(w^{(1)}_n)\) | \(\tilde{K}_n(w^{(2)}_n)\) | \(\hat{v}_n(w^{(1)}_n)\) | \(\hat{v}_n(w^{(2)}_n)\) | \(\tilde{v}_n(w^{(1)}_n)\) | \(\tilde{v}_n(w^{(2)}_n)\) |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Median         | 40.528         | 40.787         | 38.777         | 36.647         | 0.577          | 0.602          | 0.583          | 0.655          |
| Mean           | 44.870         | 42.799         | 43.078         | 37.328         | 0.755          | 0.603          | 0.583          | 0.657          |
| SD             | 36.711         | 21.896         | 36.819         | 14.729         | 0.040          | 0.027          | 0.055          | 0.037          |
| RMSE           | 0.549          | 0.212          | 0.561          | 0.151          | 0.010          | 0.016          | 0.017          | 0.051          |
| \(n = 500\)    | 49.454         | 51.329         | 49.292         | 50.315         | 0.544          | 0.552          | 0.545          | 0.575          |
| Median         | 49.788         | 51.306         | 47.570         | 50.287         | 0.544          | 0.553          | 0.544          | 0.576          |
| Mean           | 6.094          | 4.336          | 8.592          | 4.349          | 0.015          | 0.011          | 0.019          | 0.012          |
| SD             | 0.015          | 0.008          | 0.032          | 0.008          | 0.014          | 0.008          | 0.019          | 0.011          |
| RMSE           | 0.003          | 0.003          | 0.004          | 0.002          | 0.0002         | 0.0003         | 0.0003         | 0.002          |
| \(n = 2000\)   | 49.840         | 51.519         | 49.776         | 51.075         | 0.540          | 0.548          | 0.540          | 0.560          |
| Median         | 49.707         | 51.531         | 49.518         | 51.076         | 0.541          | 0.548          | 0.541          | 0.561          |
| Mean           | 2.685          | 2.086          | 3.001          | 2.120          | 0.008          | 0.006          | 0.009          | 0.005          |
| SD             | 0.003          | 0.003          | 0.004          | 0.002          | 0.0002         | 0.0003         | 0.0003         | 0.002          |
| RMSE           | 0.003          | 0.003          | 0.004          | 0.002          | 0.0002         | 0.0003         | 0.0003         | 0.002          |

studying a population because it is still alive (i.e. under the condition \(Z_n > 0\)) and the data
contains observations of small population sizes (< 10 individuals), then a \(C\)-consistent estimator
such as \(\hat{\theta}_n\) may be preferable to other similar estimators such as \(\tilde{\theta}_n\). On the other hand, if
there are no observations at small population sizes, then a \(C\)-consistent estimator such as \(\hat{\theta}_n\)
will generally be equivalent to other estimators such as \(\tilde{\theta}_n\).

6 Application to the Chatham Island black robins

We now consider the real-world application of the Chatham Island black robins who breed
seasonally and compete for territory, and for which a discrete-time PSDBP is therefore an
appropriate model. We use a model slightly different from the one considered in Section 5, due
to the characteristics of the population and the data which are available. The data consist of
yearly nesting attempts, number of offspring when a nesting attempt is successful, and bird
lifetimes between 1970 and 2000 ([@3][@12]).

More precisely, individuals in our model are birds who are at least one-year old and able
to reproduce. For every mother bird in year \(n\), the following events are assumed to happen in
this specific order between year \(n\) and year \(n + 1\):

- the mother makes a successful attempt to reproduce with probability \(r = r(z, K, v)\) which
  depends on the current population size \(z\), the carrying capacity \(K\), and an ‘efficiency’
  parameter \(v\),

- if there is a successful reproduction attempt then the mother produces a random number
  of female birds who survive to the next year with distribution \(b = (b_k, k \geq 0)\). The
  empirical distribution we obtained from the data is

\[ b = [0.3130, 0.4544, 0.1758, 0.0431, 0.0106, 0.0024, 0.0006, 0.0001], \]

which is well approximated by a binomial distribution with parameters \(n = 5\) and \(p = 0.1988\) (fitted using the method of moments).

- the mother survives to the next year with probability \(1 - d = 0.6861\) (estimated from the data), which we assume to be independent of the current population size and of whether
she had a successful reproduction attempt.
Taking into account that, if the mother survives, then she is counted among her progeny, the effective offspring distribution is \( p = (p_k), k \geq 0 \), with

\[
\begin{align*}
p_0 &= (1 - r) d + r b_0 d \\
p_1 &= (1 - r) (1 - d) + r [b_0 (1 - d) + b_1 d] \\
p_k &= r [b_{k-1} (1 - d) + b_k d], \quad k \geq 2.
\end{align*}
\]

We consider the Beverton-Holt and Ricker models where the probability \( r \) of successful reproduction attempt is given by

\[
\begin{align*}
r(z, K, v) &= \frac{v K}{(\mu - 1)z + K} \quad \text{(BH model)} \\
r(z, K, v) &= v \left( \frac{1}{\mu} \right)^{\frac{z}{K}}, \quad \text{(Ricker model)}
\end{align*}
\]

where the efficiency parameter \( v \) is the probability of a successful reproduction attempt in the absence of competition, and \( \mu := 5pv/d \). We estimate the two parameters \( v \) and \( K \) based on the yearly number of adult females between 1972 and 1998.

In Table 5 we present the estimates obtained with the estimators \( \hat{\theta}_n \) and \( \tilde{\theta}_n \) and the two weighting functions \( \{\hat{w}_n^{(1)}(z)\} \) and \( \{\hat{w}_n^{(2)}(z)\} \). In Figure 11 we plot the yearly population counts of the number of adult females and together with the mean population size curves of the Beverton Holt models and Ricker model estimated using weighting function \( \{\hat{w}_n^{(2)}(z)\} \).

From Table 5 we see that, for both estimators \( \hat{\theta}_n \) and \( \tilde{\theta}_n \), there is a large difference in the estimates obtained using the two weighting functions. This is due to the fact that the number of breeding females was one for seven years in a row in the 1970’s (as can be seen in Figure 11), and the function \( \{\hat{w}_n^{(1)}(z)\} \) places a significant amount of weight on these observations. As highlighted in Section 5.1, the weighting function \( \{\hat{w}_n^{(1)}(z)\} \) does not generally provide results as accurate as the weighting function \( \{\hat{w}_n^{(2)}(z)\} \); so here, we rely more on the results obtained with \( \{\hat{w}_n^{(2)}(z)\} \).

We also observe from Table 5 that the estimates are quite sensitive to the chosen model. This is also highlighted in Figure 11 especially for the years after 1998 where the models are used for making predictions. The black robin population reached 239 adults in 2011 (that is, approximately 120 females) \[18\], so it appears that the Beverton-Holt model is more appropriate here than the Ricker model. Several other PSDBP models could be fitted to the black robin data, and a thorough model selection analysis should then be performed, in collaboration with experts in ecology.

7 Proofs

Proof of Lemma 1. The fact that \( \hat{w}_z^{(1)} \) satisfies Assumption (A1) is a consequence of [7, Theorem 3(c)] (see also [2, Lemma 1]).

The fact that \( \hat{w}_z^{(2)} \) satisfies Assumption (A1) follows from the fact that

\[
\hat{w}_n^{(2)}(z) = \frac{z \sum_{i=0}^{n-1} 1 \{Z_i = z\} Z_k}{\sum_{k=1}^{n} Z_k} = z \frac{\hat{w}_z^{(1)}}{\sum_{k=1}^{n} Z_k/n}. \quad (21)
\]
Table 5: Black robin population. Estimates for the carrying capacity $K$ and the efficiency parameter $v$ obtained via each method.

| Method | Beverton-Holt $\hat{K}$ | Beverton-Holt $\hat{v}$ | Ricker $\hat{K}$ | Ricker $\hat{v}$ |
|--------|--------------------------|--------------------------|------------------|------------------|
| $\theta_n(w_n^{(1)})$ | 256.8276 | 0.4492 | 242.1395 | 0.4521 |
| $\hat{\theta}_n(w_n^{(2)})$ | 109.2115 | 0.6988 | 95.5616 | 0.6797 |
| $\tilde{\theta}_n(w_n^{(1)})$ | 160.8160 | 0.5857 | 129.6446 | 0.5863 |
| $\tilde{\theta}_n(w_n^{(2)})$ | 107.3594 | 0.7139 | 94.0594 | 0.6930 |

Conditional on $Z_n > 0$, the numerator of (21) converges in probability to $u_zv_z$, and the denominator converges to $\sum_{k=1}^{\infty} ku_zv_z$ in probability by [2, Lemma 6]. The result then follows from Slutsky’s theorem.

To prove Theorems 1 and 2 we require the following two lemmas.

**Lemma 2.** If the weights $\{\hat{w}_n(z)\}_{z \geq 1}$ satisfy Assumption (A1) then for each $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}_i \left[ \sum_{z=1}^{\infty} |\hat{w}_n(z) - w_z| > \epsilon |Z_n > 0 | \right] = 0.$$  

**Proof of Lemma 2.** First recall that $\{\hat{w}_n(z)\}$ and $\{w_z\}$ are probability distributions. Therefore, for any $\epsilon > 0$, we can choose $z^*(\epsilon)$ such that

$$\sum_{z=z^*(\epsilon)+1}^{\infty} w_z \leq \frac{\epsilon}{4}.$$  

Observe that if

$$|\hat{w}_n(z) - w_z| < \frac{\epsilon}{4z^*(\epsilon)}, \text{ for all } z \in \{1, \ldots, z^*(\epsilon)\},$$  

Figure 11: Black robin population. Number of adult females between 1972 and 1998 (blue points), together with the mean population size curve of the estimated Beverton Holt (red curve) and Ricker (green curve) models obtained with our estimator $\hat{\theta}_n$ and the weighting function.
then
\[
\sum_{z=1}^{\infty} |\hat{w}_n(z) - w_z| \leq \sum_{z=1}^{z^*(\varepsilon)} |\hat{w}_n(z) - w_z| + \sum_{z=z^*(\varepsilon)+1}^{\infty} w_z + \sum_{z=z^*(\varepsilon)+1}^{\infty} \hat{w}_n(z)
\]
\[
\leq \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + 1 - \sum_{z=1}^{z^*(\varepsilon)} \hat{w}_n(z)
\]
\[
\leq \frac{\varepsilon}{2} + 1 - \sum_{z=1}^{z^*(\varepsilon)} w_z + \sum_{z=1}^{z^*(\varepsilon)} |\hat{w}_n(z) - w_z|
\]
\[
\leq \varepsilon,
\]
where we have applied the triangle inequality, (22), (23), and the fact that \(\{\hat{w}_n(z)\}\) is a probability distribution. Consequently,
\[
\lim_{n \to \infty} \mathbb{P}_i \left[ \sum_{z=1}^{\infty} |\hat{w}_n(z) - w_z| > \varepsilon |Z_n > 0\right]
\]
\[
\leq \lim_{n \to \infty} \mathbb{P}_i \left[ |\hat{w}_n(z) - w_z| > \frac{\varepsilon}{4z^*(\varepsilon)}, \text{ for some } z = 1, \ldots, z^*(\varepsilon) |Z_n > 0\right]
\]
\[
\leq \sum_{i=1}^{z^*(\varepsilon)} \lim_{n \to \infty} \mathbb{P}_i \left[ |\hat{w}_n(z) - w_z| > \frac{\varepsilon}{4z^*(\varepsilon)} |Z_n > 0\right]
\]
\[
= 0,
\]
where in the final step we apply Assumption (A1).

**Lemma 3.** Let \(i \in \mathbb{N}\) be an initial state and \(f : \Theta \in \Theta \mapsto [0, \infty)\) be a continuous function with a unique minimum at \(\theta_0\) that satisfies \(f(\theta_0) = 0\). For each \(n \in \mathbb{N}\), let \(f_n : (\omega, \theta) \in \Omega \times \Theta \mapsto \mathbb{R}_+\) be a function such that for each \(\omega \in \Omega\), \(f_n(\omega, \cdot)\) is a continuous function on \(\Theta\), and for each \(\theta \in \Theta\), \(f_n(\cdot, \theta)\) is a measurable function of the random variables \(Z_0, \ldots, Z_n\). If
\[
\forall \varepsilon > 0, \lim_{n \to \infty} \mathbb{P}_i \left[ \omega \in \Omega : \sup_{\theta \in \Theta} |f_n(\omega, \theta) - f(\theta)| > \varepsilon |Z_n > 0\right] = 0,
\]
then

(i) \(\lim_{n \to \infty} \mathbb{P}_i [\omega \in \Omega : \exists \arg\min_{\theta \in \Theta} f_n(\omega, \theta) |Z_n > 0] = 1\).

(ii) \(\forall \varepsilon > 0, \lim_{n \to \infty} \mathbb{P}_i [\omega \in \Omega : \| \arg\min_{\theta \in \Theta} f_n(\omega, \theta) - \theta_0\| > \varepsilon |Z_n > 0] = 0\).

**Proof.** To ease the notation, throughout this proof let us denote \(f_n(\theta) = f_n(\cdot, \theta)\). To establish (i) and (ii) we apply the following two identities that hold under (24): for any \(\varepsilon > 0\), and \(C \subseteq \Theta\) we have
\[
\lim_{n \to \infty} \mathbb{P}_i [f_n(\theta_0) > \varepsilon |Z_n > 0] = 0,
\]
\[
\lim_{n \to \infty} \mathbb{P}_i \left[ \inf_{\theta \in C} f_n(\theta) - \inf_{\theta \in C} f(\theta) > \varepsilon |Z_n > 0\right] = 0.
\]
To establish (25) we use the fact that, since \( f(\theta_0) = 0 \), for any \( n \in \mathbb{N} \) we have
\[
\{ \omega \in \Omega : f_n(\theta_0) > \varepsilon \} = \{ \omega \in \Omega : |f_n(\theta_0) - f(\theta_0)| > \varepsilon \} \subseteq \{ \omega \in \Omega : \sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)| > \varepsilon \},
\]
which, by (24), implies
\[
\lim_{n \to \infty} \mathbb{P}_i \left[ f_n(\theta_0) > \varepsilon | Z_n > 0 \right] \leq \lim_{n \to \infty} \mathbb{P}_i \left[ \sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)| > \varepsilon | Z_n > 0 \right] = 0.
\]
To establish (26), we denote \( \theta_C = \arg \inf_{\theta \in C} f(\theta) \) and \( \theta_{n,C} = \arg \inf_{\theta \in C} f_n(\theta) \). For any \( \theta' \in C \) we have (surely)
\[
\inf_{\theta \in C} f_n(\theta) - \inf_{\theta \in C} f(\theta) \leq f_n(\theta') - \inf_{\theta \in C} f(\theta),
\]
and consequently,
\[
\inf_{\theta \in C} f_n(\theta) - \inf_{\theta \in C} f(\theta) \leq f_n(\theta_C) - \inf_{\theta \in C} f(\theta) = f_n(\theta_C) - f(\theta_C) \leq \sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)|.
\]
Similarly,
\[
\inf_{\theta \in C} f(\theta) - \inf_{\theta \in C} f_n(\theta) \leq f(\theta_n) - \inf_{\theta \in C} f_n(\theta) = f(\theta_n) - f_n(\theta_n) \leq \sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)|.
\]
Combining both the above inequalities we see that for any \( n \in \mathbb{N} \) and \( \varepsilon > 0 \),
\[
\left\{ \left| \inf_{\theta \in C} f_n(\theta) - \inf_{\theta \in C} f(\theta) \right| > \varepsilon \right\} \subseteq \left\{ \sup_{\theta \in \Theta} |f_n(\theta) - f(\theta)| > \varepsilon \right\},
\]
which then allows us to apply (24) to obtain (26).

To establish (i) we observe that it is sufficient to show that, for any compact set \( C \subseteq \Theta \) with \( \theta_0 \in C \), we have
\[
\lim_{n \to \infty} \mathbb{P}_i \left[ \inf_{\theta \in \Theta \setminus C} f_n(\theta) > \min_{\theta \in C} f_n(\theta) | Z_n > 0 \right] = 1. \tag{27}
\]
Note that, because \( C \) is a compact set, and because for any \( n \in \mathbb{N} \) and \( \omega \in \Omega, \theta \mapsto f_n(\omega, \theta) \) is a continuous function on \( \Theta \), \( \min_{\theta \in C} f_n(\theta) \) exists. The claim in (i) then follows from (27) by letting \( C \) be an arbitrarily small closed ball centered at \( \theta_0 \).

To establish (27) we let \( \varepsilon = a/4 \), with \( a = \inf_{\theta \in \Theta \setminus C} f(\theta) > 0 \), and observe that
\[
\left\{ f_n(\theta_0) \leq \varepsilon \right\} \cap \left\{ \left| \inf_{\theta \in \Theta \setminus C} f_n(\theta) - a \right| \leq \varepsilon \right\} \subseteq \left\{ \min_{\theta \in C} f_n(\theta) \leq f_n(\theta_0) \leq \frac{a}{4} < \frac{3a}{4} \leq \inf_{\theta \in \Theta \setminus C} f_n(\theta) \right\}.
\]
We then apply (25) and (26) to obtain (27).

To establish (ii) we let \( \vartheta_n = \arg \min_{\theta \in \Theta} f_n(\theta) \), and use the fact that
\[
\forall \varepsilon > 0, \lim_{n \to \infty} \mathbb{P}_i \left[ f(\vartheta_n) > \varepsilon | Z_n > 0 \right] = 0. \tag{28}
\]
Indeed, (28) follows by applying both (26) with $C = \Theta$, to obtain
\[ \forall \epsilon > 0, \lim_{n \to \infty} \mathbb{P}_i \left[ f_n(\theta_n) > \epsilon | Z_n > 0 \right] = 0, \]
and (24) to obtain
\[ \forall \epsilon > 0, \lim_{n \to \infty} \mathbb{P}_i \left[ | f_n(\theta_n) - f(\theta_n) | > \epsilon | Z_n > 0 \right] = 0. \]

To prove (ii), we then note that, for each $\epsilon > 0$, given that $f(\cdot)$ has a unique minimum at $\theta_0$, we can take $C = \{ \theta \in \Theta : \| \theta - \theta_0 \| \leq \epsilon \}$, $b = \inf_{\Theta \cap C} f(\theta)$, and $\beta = b/2$; then $0 = \min_{\theta \in C} f(\theta) < \beta < b$. Moreover, we have $b = \inf_{\Theta \setminus C} f(\theta) \leq f(\theta')$ for every $\theta' \in \Theta \setminus C$, and consequently
\[ \{ \| \theta_n - \theta_0 \| > \epsilon \} \subseteq \{ f(\theta_n) > \beta \}. \]
The result in (ii) now follows from (28). \hfill \square

We are now ready to establish the limiting properties of the estimator $\hat{\theta}_n$, namely, $C$-consistency (Theorem 1) and $C$-normality (Theorem 2). For each $n \in \mathbb{N}$, recall that
\[ \hat{\theta}_n = \arg \min_{\theta \in \Theta} S_n(\theta), \quad \text{where} \quad S_n(\theta) := \sum_{z=1}^{\infty} \hat{w}_n(z)(m^\uparrow(z, \theta) - \hat{m}_n(z))^2. \]

Proof of Theorem 1
Let us fix $i \in \mathbb{N}$. By Lemma 3, it is enough to prove that
\[ \forall \epsilon > 0, \lim_{n \to \infty} \mathbb{P}_i \left[ \sup_{\theta \in \Theta} \left| S_n(\theta) - S(\theta) \right| > \epsilon | Z_n > 0 \right] = 0, \]
where
\[ S(\theta) := \sum_{z=1}^{\infty} w_z(\theta_0)(m^\uparrow(z, \theta) - m^\uparrow(z, \theta_0))^2. \]

To that end, we introduce the following function
\[ \hat{S}_n(\theta) = \sum_{z=1}^{\infty} w_z(\theta_0)1_{\{j_n(z) > 0\}}(m^\uparrow(z, \theta) - \hat{m}_n(z))^2, \quad n \in \mathbb{N}, \theta \in \Theta, \]
where $j_n(z) := \sum_{i=0}^{n-1} 1_{\{Z_i = z\}}$. It then suffices to prove that, for each $\epsilon > 0$,
\[ (a) \lim_{n \to \infty} \mathbb{P}_i \left[ \sup_{\theta \in \Theta} \left| S_n(\theta) - \hat{S}_n(\theta) \right| > \epsilon | Z_n > 0 \right] = 0, \]
\[ (b) \lim_{n \to \infty} \mathbb{P}_i \left[ \sup_{\theta \in \Theta} \left| \hat{S}_n(\theta) - S(\theta) \right| > \epsilon | Z_n > 0 \right] = 0. \]

We start with the proof of (a). Recall the definition of $M$ and $C$ from Assumptions (A3) and (A4) respectively. Let $M' = 2(M + C)^2$ and $A_n = \sup_{z \in \mathbb{N}} |m^\uparrow(z, \theta_0) - \hat{m}_n(z)|$. On the set $\{ A_n \leq C \}$ we have
\[ \sup_{\theta \in \Theta} |S_n(\theta) - \hat{S}_n(\theta)| \leq \sup_{\theta \in \Theta} \sum_{z=1}^{\infty} \left| \hat{w}_n(z) - w_z(\theta_0)1_{\{j_n(z) > 0\}} \right| (m^\uparrow(z, \theta) - \hat{m}_n(z))^2 \]
\[
\sup_{\theta \in \Theta} \sum_{z=1}^{\infty} |\hat{w}_n(z) - w_z(\theta_0)1_{\{j_n(z) > 0\}}| (m^\top(z, \theta) - m^\top(z, \theta_0) + m^\top(z, \theta_0) - \hat{m}_n(z))^2 \\
\leq \sup_{\theta \in \Theta} \sum_{z=1}^{\infty} |\hat{w}_n(z) - w_z(\theta_0)1_{\{j_n(z) > 0\}}| (m^\top(z, \theta) - m^\top(z, \theta_0))^2 \\
+ \sum_{z=1}^{\infty} |\hat{w}_n(z) - w_z(\theta_0)1_{\{j_n(z) > 0\}}| (m^\top(z, \theta_0) - \hat{m}_n(z))^2 \\
+ 2 \sup_{\theta \in \Theta} \sum_{z=1}^{\infty} |\hat{w}_n(z) - w_z(\theta_0)1_{\{j_n(z) > 0\}}| \cdot |m^\top(z, \theta) - m^\top(z, \theta_0)| \cdot |m^\top(z, \theta_0) - \hat{m}_n(z)| \\
\leq M' \sum_{z=1}^{\infty} |\hat{w}_n(z) - w_z(\theta_0)1_{\{j_n(z) > 0\}}| \\
\leq M' \sum_{z=1}^{\infty} |\hat{w}_n(z) - w_z(\theta_0) + w_z(\theta_0) - w_z(\theta_0)1_{\{j_n(z) > 0\}}| \\
\leq M' \sum_{z=1}^{\infty} |\hat{w}_n(z) - w_z(\theta_0)| + M' \sum_{z=1}^{\infty} w_z(\theta_0)1_{\{j_n(z) = 0\}}.
\]

Consequently,
\[
\left\{ \sup_{\theta \in \Theta} |S_n(\theta) - \hat{S}_n(\theta)| > \epsilon \right\} = \left\{ \sup_{\theta \in \Theta} |S_n(\theta) - \hat{S}_n(\theta)| > \epsilon, A_n \leq C \right\} \cup \\
\bigcup \left\{ \sup_{\theta \in \Theta} |S_n(\theta) - \hat{S}_n(\theta)| > \epsilon, A_n > C \right\} \\
\subseteq \left\{ \sum_{z=1}^{\infty} |\hat{w}_n(z) - w_z(\theta_0)| > \frac{\epsilon}{2M'}, A_n \leq C \right\} \cup \\
\bigcup \left\{ \sum_{z=1}^{\infty} w_z(\theta_0)1_{\{j_n(z) = 0\}} > \frac{\epsilon}{2M'}, A_n \leq C \right\} \cup \{A_n > C\} \\
\subseteq \left\{ \sum_{z=1}^{\infty} |\hat{w}_n(z) - w_z(\theta_0)| > \frac{\epsilon}{2M'} \right\} \cup \left\{ \sum_{z=1}^{\infty} w_z(\theta_0)1_{\{j_n(z) = 0\}} > \frac{\epsilon}{2M'} \right\} \\
\cup \{A_n > C\}.
\]

Conditioning on \(Z_n > 0\), we then obtain \(a\) by applying Lemma 2 to the first event, the fact that \(\lim_{n \to \infty} P[1_{\{j_n(z) = 0\}} \leq \epsilon | Z_n > 0| = 1\) for any \(\epsilon > 0, z \in \mathbb{N}\) and that \(\sum_{z=1}^{\infty} w_z(\theta_0) = 1\) to the second event, and Assumption \(A4\) to the third event.

In order to prove \(b\) we observe that
\[
\hat{S}_n(\theta) - S(\theta) = \sum_{z=1}^{\infty} w_z(\theta_0)1_{\{j_n(z) > 0\}} \left[ (m^\top(z, \theta) - \hat{m}_n(z))^2 - (m^\top(z, \theta) - m^\top(z, \theta_0))^2 \right] \\
- \sum_{z=1}^{\infty} w_z(\theta_0)1_{\{j_n(z) = 0\}} (m^\top(z, \theta) - m^\top(z, \theta_0))^2
\]
\[ \begin{align*}
&= \sum_{z=1}^{\infty} w_z(\theta_0) \mathbb{1}_{\{s_n(z) > 0\}} \left[ \hat{m}_n(z)^2 - 2\hat{m}_n(z)m^\top(z, \theta) - m^\top(z, \theta_0)^2 + 2m^\top(z, \theta)m^\top(z, \theta_0) \right] \\
&- \sum_{z=1}^{\infty} w_z(\theta_0) \mathbb{1}_{\{s_n(z) = 0\}} (m^\top(z, \theta) - m^\top(z, \theta_0))^2,
\end{align*} \]

then

\[ \sup_{\theta \in \Theta} |\hat{S}_n(\theta) - S(\theta)| \leq \sum_{z=1}^{\infty} w_z(\theta_0) \mathbb{1}_{\{s_n(z) > 0\}} |\hat{m}_n(z)^2 - m^\top(z, \theta_0)^2| \\
+ 2 \sup_{\theta \in \Theta} \sum_{z=1}^{\infty} w_z(\theta_0) \mathbb{1}_{\{s_n(z) > 0\}} m^\top(z, \theta) |\hat{m}_n(z) - m^\top(z, \theta_0)| \\
+ \sup_{\theta \in \Theta} \sum_{z=1}^{\infty} w_z(\theta_0) \mathbb{1}_{\{s_n(z) = 0\}} (m^\top(z, \theta) - m^\top(z, \theta_0))^2, \]

and consequently,

\[ \left\{ \sup_{\theta \in \Theta} |\hat{S}_n(\theta) - S(\theta)| > \epsilon \right\} \subseteq \left\{ \sum_{z=1}^{\infty} w_z(\theta_0) \mathbb{1}_{\{s_n(z) > 0\}} |\hat{m}_n(z)^2 - m^\top(z, \theta_0)^2| > \frac{\epsilon}{3} \right\} \cup \quad (29) \]

\[ \cup \left\{ \sup_{\theta \in \Theta} \sum_{z=1}^{\infty} w_z(\theta_0) \mathbb{1}_{\{s_n(z) > 0\}} m^\top(z, \theta) |\hat{m}_n(z) - m^\top(z, \theta_0)| > \frac{\epsilon}{6} \right\} \quad (30) \]

\[ \cup \left\{ \sup_{\theta \in \Theta} \sum_{z=1}^{\infty} w_z(\theta_0) \mathbb{1}_{\{s_n(z) = 0\}} (m^\top(z, \theta) - m^\top(z, \theta_0))^2 > \frac{\epsilon}{3} \right\} \quad (31) \]

Because \( \mathbb{P}_1(A_n > C|Z_n > 0) \to 0 \) by (A4), we can restrict our attention to the set \( \{A_n \leq C\} \).

On the set \( \{A_n \leq C\} \),

\[ \sup_{z \in \mathbb{N}} |\hat{m}_n(z)^2 - m^\top(z, \theta_0)^2| \leq \sup_{z \in \mathbb{N}} |\hat{m}_n(z) - m^\top(z, \theta_0)| \cdot \sup_{z \in \mathbb{N}} |\hat{m}_n(z) + m^\top(z, \theta_0)| \]

\[ \leq C(C + 2M) \leq M'. \]

In addition, since \( \mathbf{w}(\theta_0) \) is a probability distribution, given \( \epsilon > 0 \), there exists \( K_0 = K_0(\epsilon) \in \mathbb{N} \) such that \( \sum_{z=K_0+1}^{\infty} w_z(\theta_0) < \frac{\epsilon}{6M'} \). Consequently, on \( \{A_n \leq C\} \)

\[ \sum_{z=1}^{\infty} w_z(\theta_0) \mathbb{1}_{\{s_n(z) > 0\}} |\hat{m}_n(z)^2 - m^\top(z, \theta_0)^2| \leq \sum_{z=1}^{K_0} w_z(\theta_0) \mathbb{1}_{\{s_n(z) > 0\}} |\hat{m}_n(z)^2 - m^\top(z, \theta_0)^2| \\
+ M' \sum_{z=K_0+1}^{\infty} w_z(\theta_0), \]

and hence

\[ \bigcap_{z=1}^{K_0} \left\{ w_z(\theta_0) \mathbb{1}_{\{s_n(z) > 0\}} |\hat{m}_n(z)^2 - m^\top(z, \theta_0)^2| \leq \frac{\epsilon}{6K_0} \right\} \bigcap \{A_n \leq C\} \subseteq \]

25
Thus, by (15), we get that
\[ \lim_{n \to \infty} \mathbb{P}_i \left[ \left\{ \sum_{z=1}^{\infty} w_z(\theta_0) \mathbf{1}_{\{j_n(z) > 0\}} \left| \hat{m}_n(z)^2 - m^\dagger(z, \theta_0)^2 \right| \leq \frac{\epsilon}{3} \right\} \cap \{A_n \leq C\} \right] Z_n > 0 \right] = 1, \]
and then
\[ \lim_{n \to \infty} \mathbb{P}_i \left[ \left\{ \sum_{z=1}^{\infty} w_z(\theta_0) \mathbf{1}_{\{j_n(z) > 0\}} \left| \hat{m}_n(z)^2 - m^\dagger(z, \theta_0)^2 \right| \leq \frac{\epsilon}{3} \right\} \cap \{A_n \leq C\} \right] Z_n > 0 \right] = 1. \]

With similar arguments using \( \sup_{\theta \in \Theta} \sup_{z \in \mathbb{N}} (m^\dagger(z, \theta) - m^\dagger(z, \theta_0))^2 \leq 4M^2 \), and we obtain
\[ \lim_{n \to \infty} \mathbb{P}_i \left[ \sup_{\theta \in \Theta} \left\{ \sum_{z=1}^{\infty} w_z(\theta_0) \mathbf{1}_{\{j_n(z) > 0\}} m^\dagger(z, \theta) \left| \hat{m}_n(z) - m^\dagger(z, \theta_0) \right| \leq \frac{\epsilon}{6} \right\} \cap \{A_n \leq C\} \right] Z_n > 0 \right] = 1, \]
\[ \lim_{n \to \infty} \mathbb{P}_i \left[ \sup_{\theta \in \Theta} \left\{ \sum_{z=1}^{\infty} w_z(\theta_0) \mathbf{1}_{\{j_n(z) = 0\}} (m^\dagger(z, \theta) - m^\dagger(z, \theta_0))^2 \leq \frac{\epsilon}{3} \right\} \cap \{A_n \leq C\} \right] Z_n > 0 \right] = 1. \]

Consequently we have proved (b) and the result now follows.

Before we prove Theorem 2 we recall the asymptotic normality of the MLEs \( \{\hat{m}_n(z)\}_{z \in \mathbb{N}} \) that is established in [22, Theorem 1]: for any \( i, z, z_1, z_2 \in \mathbb{N} \) and \( x, x_1, x_2 \in \mathbb{R} \),
\[ \lim_{n \to \infty} \mathbb{P}_i [\sqrt{n}(\hat{m}_n(z) - m^\dagger(z, \theta_0)) \leq x | Z_n > 0] = \Psi_{\gamma(z, \theta_0)}(x), \tag{32} \]
\[ \lim_{n \to \infty} \mathbb{P}_i [\sqrt{n}(\hat{m}_n(z_i) - m^\dagger(z_i, \theta_0)) \leq x_i, \ i = 1, 2 | Z_n > 0] = \Psi_{\Gamma(z_1, z_2, \theta_0)}(x_1, x_2), \tag{33} \]
where \( \gamma(z, \theta_0) \) is defined in (19) and
\[ \Gamma(z_1, z_2, \theta_0) := \begin{pmatrix} \gamma(z_1, \theta_0) & 0 \\ 0 & \gamma(z_2, \theta_0) \end{pmatrix}. \tag{34} \]

**Proof of Theorem 2**

Let us fix \( i \in \mathbb{N} \). First, we use a Taylor expansion for the function \( \nabla S_n(\cdot) \) around \( \theta_0 \):
\[ 0 = \nabla S_n(\hat{\theta}_n) = \nabla S_n(\theta_0) + \nabla^2 S_n(\theta_0) \left( \hat{\theta}_n - \theta_0 \right), \]
surely, where \( \theta_n \) is a point between \( \hat{\theta}_n \) and \( \theta_0 \), and \( \nabla^2 S_n(\theta_n) \) is the Jacobian of \( S_n(\cdot) \) at \( \theta_n \). Then,
\[ \sqrt{n}(\hat{\theta}_n - \theta_0) = - (\nabla^2 S_n(\theta_n))^{-1} \sqrt{n} \nabla S_n(\theta_0), \tag{35} \]
where we used the fact that \( \nabla^2 S_n(\theta_n) \) is a symmetric matrix. By Assumption (A5), we can interchange derivatives and infinite sum in the function \( S_n(\cdot) \). Given Equation (35), by Slutsky’s theorem, it is then sufficient to establish (a) and (b) below:
(a) If $\Psi_{\zeta(\theta_0)}(\cdot)$ is the distribution function of a $d$-dimensional normal distribution with mean vector $0 = (0, \ldots, 0)^\top$ and covariance matrix $\zeta(\theta_0)$ and $x_1, \ldots, x_d \in \mathbb{R}$, then
\[
\lim_{n \to \infty} \mathbb{P}_i \left[ -\sqrt{n} \frac{\partial S_n(\theta_0)}{\partial \theta_j} \leq x_j, \ j = 1, \ldots, d \mid Z_n > 0 \right] = \Psi_{\zeta(\theta_0)}(x_1, \ldots, x_d).
\]

(b) For each $\epsilon > 0$,
\[
\lim_{n \to \infty} \mathbb{P}_i \left[ \left| \frac{\partial^2 S_n(\theta_0)}{\partial \theta_j \partial \theta_l} - \eta_{jl}(\theta_0) \right| > \epsilon \mid Z_n > 0 \right] = 0, \ j, l = 1, \ldots, d.
\]

To prove (a) we fix $j = 1, \ldots, d$ and denote
\[
U_{nk} = \sum_{z=1}^{k} \frac{\partial m^\dagger(z, \theta_0)}{\partial \theta_j} \hat{w}_n(z) \sqrt{n} (m^\dagger(z, \theta_0) - \hat{m}_n(z)), \ n, k \in \mathbb{N},
\]
\[
\rho(k, \theta_0) = \sum_{z=1}^{k} w_z(\theta_0)^2 \left( \frac{\partial m^\dagger(z, \theta_0)}{\partial \theta_j} \right)^2 \gamma(z, \theta_0), \ k \in \mathbb{N},
\]
where we recall the definition of $\gamma(z, \theta_0)$ from (19). We then apply [1, Theorem 25.5] which states that, for a two-dimensional random vector $(X^{(k)}_n, Y_n)$ satisfying
\begin{enumerate}[(i)]
  \item for each $k$, $X^{(k)}_n \xrightarrow{d} X$ as $n \to \infty$,
  \item $X^{(k)} \xrightarrow{d} X$ as $k \to \infty$, and
  \item $\lim_{k \to \infty} \limsup_{n \to \infty} \mathbb{P}[|X^{(k)}_n - Y_n| \geq \varepsilon] = 0$ for all $\varepsilon > 0$,
\end{enumerate}
we have $Y_n \xrightarrow{d} X$, where in our case $X^{(k)}_n = U_{nk}$ and $Y_n = \sqrt{n} \frac{\partial S_n(\theta_0)}{\partial \theta_j}$. We now establish (i)–(iii).

(a-i) For each $k \in \mathbb{N}$ fixed we have
\[
\lim_{n \to \infty} \mathbb{P}_i[U_{nk} \leq x | Z_n > 0] = \Phi_{\rho(k, \theta_0)}(x), \ x \in \mathbb{R},
\]
where $\Phi_{\rho(k, \theta_0)}(\cdot)$ is the distribution function of a normal distribution with mean equal to 0 and variance $\rho(k, \theta_0)$. Indeed, this follows by applying (32) and (33) in combination with Assumption (A1) and Slutsky's theorem.

(a-ii) Let $U_k$ be a normal distribution with mean 0 and variance $\rho(k, \theta_0)$. Then, as $k \to \infty$, $U_k$ converges in distribution to a normal distribution with mean 0 and variance $\lim_{k \to \infty} \rho(k, \theta_0) = \frac{\zeta_{jj}(\theta_0)}{4}$.

(a-iii) It remains to show that, for each $\epsilon > 0$,
\[
\lim_{k \to \infty} \left( \limsup_{n \to \infty} \mathbb{P}_i \left[ \left| \sum_{z=1}^{\infty} \frac{\partial m^\dagger(z, \theta_0)}{\partial \theta_j} \hat{w}_n(z) \sqrt{n} (m^\dagger(z, \theta_0) - \hat{m}_n(z)) - U_{nk} \right| > \epsilon | Z_n > 0 \right] \right) = 0,
\]

27
or equivalently
\[
\lim_{k \to \infty} \left( \limsup_{n \to \infty} \mathbb{P}_i \left[ \sum_{z=k+1}^{\infty} \frac{\partial m^*(z, \theta_0)}{\partial \theta_i} \hat{w}_n(z) \sqrt{n}(m^*(z, \theta_0) - \hat{m}_n(z)) > \epsilon | Z_n > 0 \right] \right) = 0.
\]

In the case where there exists a value \( z^* < \infty \) such that for all \( n \in \mathbb{N} \) and \( z \geq z^* \), \( \hat{w}_n(z) = 0 \), (a-iii) is immediate. When \( \hat{w}_n(z) = \hat{w}_n^{(1)}(z) \) or \( \hat{w}_n(z) = \hat{w}_n^{(2)}(z) \), (a-iii) is more technically demanding, and consequently we postpone the proof of (a-iii) in that case until Lemma 4 in Appendix A.

Finally, in order to prove (b) let us fix \( j, l = 1, \ldots, d \), and write
\[
T_n^{(1)} = \sum_{z=1}^{\infty} \frac{\partial^2 m^*(z, \theta_n)}{\partial \theta_j \partial \theta_l} \hat{w}_n(z) (m^*(z, \theta_n) - \hat{m}_n(z)), \quad n \in \mathbb{N}
\]
\[
T_n^{(2)} = \sum_{z=1}^{\infty} \frac{\partial m^*(z, \theta_n)}{\partial \theta_j} \frac{\partial m^*(z, \theta_n)}{\partial \theta_l} \hat{w}_n(z) - \frac{\eta l_j(\theta_0)}{2}, \quad n \in \mathbb{N}.
\]

Because \( \frac{\partial^2 S_n(\theta_n)}{\partial \theta_i \partial \theta_j} - \eta l_j(\theta_0) = 2(T_n^{(1)} + T_n^{(2)}) \), by the triangle inequality, it suffices to prove that, for each \( \epsilon > 0 \),
\[
(b-1) \quad \lim_{n \to \infty} \mathbb{P}_i \left[ |T_n^{(1)}| > \epsilon | Z_n > 0 \right] = 0.
\]
\[
(b-2) \quad \lim_{n \to \infty} \mathbb{P}_i \left[ |T_n^{(2)}| > \epsilon | Z_n > 0 \right] = 0.
\]

Let us start with (b-1). We write
\[
T_{1n}^{(1)} = \sum_{z=1}^{\infty} \left| \frac{\partial^2 m^*(z, \theta_n)}{\partial \theta_j \partial \theta_l} \right| \cdot |m^*(z, \theta_n) - \hat{m}_n(z)| \cdot |\hat{w}_n(z) - w_z(\theta_0)|,
\]
\[
T_{2n}^{(1)} = \sum_{z=1}^{\infty} \left| \frac{\partial^2 m^*(z, \theta_n)}{\partial \theta_i \partial \theta_j} \right| \cdot |m^*(z, \theta_n) - \hat{m}_n(z)| w_z(\theta_0).
\]

Since \( |T_n^{(1)}| \leq T_{1n}^{(1)} + T_{2n}^{(1)} \), then
\[
\left\{ T_{1n}^{(1)} \leq \frac{\epsilon}{2} \right\} \cap \left\{ T_{2n}^{(1)} \leq \frac{\epsilon}{2} \right\} \subseteq \left\{ |T_n^{(1)}| \leq \epsilon \right\},
\]
and thus, it suffices to prove that \( \lim_{n \to \infty} \mathbb{P}_i \left[ T_{1n}^{(1)} \leq \epsilon/2 | Z_n > 0 \right] = 1 \), for \( \ell = 1, 2 \).

Recall the definition of \( C \) from Assumption (A4) and \( C \equiv C(\theta_0) \) from Assumption (A5). As in the proof of Theorem 1 let us write \( A_n = \sup_{z \in \mathbb{N}} |m^*(z, \theta_0) - \hat{m}_n(z)| \), and take \( \delta_1 > 0 \) sufficiently small so that \( \{ \theta \in \Theta : \|\theta - \theta_0\| \leq \delta_1 \} \subseteq C \). On the one hand, if we define the event \( B_{1n} := \{ \|\theta_n - \theta_0\| \leq \delta_1 \} \cap \{ A_n \leq C \} \), then
\[
\lim_{n \to \infty} \mathbb{P}_i [B_{1n} | Z_n > 0] = 1, \quad (36)
\]
where we have used Assumption [(A4)] and Theorem [1] in combination with the fact that 
\[ \theta_n = (1 - \lambda) \hat{\theta}'_n + \lambda \theta_0, \]
for some \(0 < \lambda < 1\), thus,
\[ \|\theta_n - \theta_0\| \leq (1 - \lambda)\|\hat{\theta}'_n - \theta_0\| \leq \|\hat{\theta}'_n - \theta_0\|. \]

On the other hand, by Assumption [(A5)-(b)] and the triangle inequality, on \(B_{1n}\) we have
\[
T^{(1)}_{1n} \leq M^* \sum_{z=1}^{\infty} (|m^\uparrow(z, \theta_n) - m^\uparrow(z, \theta_0)| + |m^\uparrow(z, \theta_0) - \hat{m}_n(z)| \|\hat{w}_n(z) - w(z, \theta_0)\|)
\leq M^*(M + C) \sum_{z=1}^{\infty} \|\hat{w}_n(z) - w(z, \theta_0)\|,
\]
where in the second inequality we used Assumptions [(A3)] and [(A4)]. Therefore, by Assumption [(A1)] and Lemma [2]
\[
\lim_{n \to \infty} \mathbb{P}_i \left[ \left\{ T^{(1)}_{1n} > \frac{\epsilon}{2} \right\} \cap B_{1n} | Z_n > 0 \right] \leq \lim_{n \to \infty} \mathbb{P}_i \left[ \sum_{z=1}^{\infty} \|\hat{w}_n(z) - w(z, \theta_0)\| > \frac{\epsilon}{2M^*(M + C)} \bigg| Z_n > 0 \right] = 0.
\]

Now, using [36], we get \(\lim_{n \to \infty} \mathbb{P}_i \left[ T^{(1)}_{1n} \leq \epsilon/2 \bigg| Z_n > 0 \right] = 1\).

On the other hand, for \(T^{(1)}_{2n}\), we note that since \(w(\theta_0)\) is a probability distribution we can select \(K_1 = K_1(\epsilon) \in \mathbb{N}\) such that
\[
\sum_{z=K_1+1}^{\infty} w(z, \theta_0) \leq \frac{\epsilon}{6M^*(M + C)}.
\]
Moreover, by Assumption [(A5)-(a)] we can choose \(0 < \delta_2 = \delta_2(\epsilon) < \delta_1\) such that on the event \(B_{2n} = \{\|\theta_n - \theta_0\| \leq \delta_2\}\) we have
\[
\max_{z=1, \ldots, K_1} |m^\uparrow(z, \theta_n) - m^\uparrow(z, \theta_0)| \leq \frac{\epsilon}{6M^*K_1},
\]
where we observe that by the same arguments as for [36],
\[
\lim_{n \to \infty} \mathbb{P}_i[B_{2n} | Z_n > 0] = 1.
\] (37)

In addition, on \(\{A_n \leq C\}\) and \(B_{2n}\), we have
\[
T^{(1)}_{2n} \leq M^* \sum_{z=1}^{K_1} |m^\uparrow(z, \theta_0) - \hat{m}_n(z)| + M^* \sum_{z=1}^{K_1} |m^\uparrow(z, \theta_n) - m^\uparrow(z, \theta_0)|
+ M^*(M + C) \sum_{z=K_1+1}^{\infty} w(z, \theta_0)
\leq M^* \sum_{z=1}^{K_1} |m^\uparrow(z, \theta_0) - \hat{m}_n(z)| + \frac{\epsilon}{6} + \frac{\epsilon}{6}
\]
As a consequence, we obtain

\[
\left\{ T_{2n}^{(1)} > \frac{\epsilon}{2} \right\} \cap B_{2n} \cap \{ A_n \leq C \} \\
\subseteq \left\{ \exists z \in \{1, \ldots, K_1\} : |m(z, \theta_0) - \hat{m}_n(z)| > \frac{\epsilon}{6M^2_1K_1} \right\} \cap B_{2n} \cap \{ A_n \leq C \}.
\]

Then, by Assumption (A4) and Theorem 1, together with (15), we obtain

\[
\lim_{n \to \infty} \mathbb{P}_i \left[ \left\{ T_{2n}^{(1)} > \frac{\epsilon}{2} \right\} \cap B_{2n} \cap \{ A_n \leq C \} | Z_n > 0 \right] = 0.
\]

This yields \( \lim_{n \to \infty} \mathbb{P}_i \left[ T_{2n}^{(1)} \leq \epsilon/2 | Z_n > 0 \right] = 1. \)

Now, to prove (b-2) we observe that \( |T_n^{(2)}| \leq T_n^{(2)} + T_{2n}^{(2)} \), with

\[
T_{1n}^{(2)} = \sum_{z=1}^{\infty} \left| \frac{\partial m^\dagger(z, \theta_n)}{\partial \theta_j} \cdot \frac{\partial m^\dagger(z, \theta_n)}{\partial \theta_i} \right| |\hat{w}_n(z) - w_z(\theta_0)|,
\]

\[
T_{2n}^{(2)} = \sum_{z=1}^{\infty} \left| \frac{\partial m^\dagger(z, \theta_n)}{\partial \theta_j} \cdot \frac{\partial m^\dagger(z, \theta_n)}{\partial \theta_i} - \frac{\partial m^\dagger(z, \theta_0)}{\partial \theta_j} \cdot \frac{\partial m^\dagger(z, \theta_0)}{\partial \theta_i} \right| w_z(\theta_0),
\]

hence, since \( \left\{ T_{1n}^{(2)} \leq \frac{\epsilon}{2} \right\} \cap \left\{ T_{2n}^{(2)} \leq \frac{\epsilon}{2} \right\} \subseteq \left\{ |T_n^{(2)}| \leq \epsilon \right\} \), it is enough to prove that

\[
\lim_{n \to \infty} \mathbb{P}_i \left[ T_{\ell n}^{(2)} \leq \frac{\epsilon}{2} | Z_n > 0 \right] = 1, \quad \text{for } \ell = 1, 2.
\]

For \( T_{1n}^{(2)} \), by Assumption (A5)(a) and the fact that on \( B_{2n} \) we have \( \theta_n \in \mathcal{C} \),

\[
B_{2n} \cap \left\{ \sum_{z=1}^{\infty} |\hat{w}_n(z) - w_z(\theta_0)| \leq \frac{\epsilon}{2M^2_1} \right\} \subseteq \left\{ T_{1n}^{(2)} \leq \frac{\epsilon}{2} \right\},
\]

which, combined with (37) and Lemma 2 leads to (38) for \( \ell = 1 \).

For \( T_{2n}^{(2)} \), we take \( K_2 = K_2(\epsilon) \in \mathbb{N} \) such that

\[
\sum_{z=K_2+1}^{\infty} w_z(\theta_0) \leq \frac{\epsilon}{8M^2_1}.
\]

Since, for each \( j = 1, \ldots, d \) and \( z \in \mathbb{N} \) fixed, the function \( \theta \in \Theta \mapsto \frac{\partial m^\dagger(z, \theta)}{\partial \theta_j} \) is continuous at \( \theta_0 \) by Assumption (A5)(a) we can choose \( \delta_3 = \delta_3(\epsilon) > 0 \) such that on \( B_{3n} = \| \theta_n - \theta_0 \| \leq \delta_3 \),

\[
\max_{z=1, \ldots, K_2} \left| \frac{\partial m^\dagger(z, \theta_n)}{\partial \theta_j} \cdot \frac{\partial m^\dagger(z, \theta_n)}{\partial \theta_i} - \frac{\partial m^\dagger(z, \theta_0)}{\partial \theta_j} \cdot \frac{\partial m^\dagger(z, \theta_0)}{\partial \theta_i} \right| \leq \frac{\epsilon}{4K_2}.
\]

On \( B_{3n} \), we then have

\[
T_{2n}^{(2)} = \sum_{z=1}^{K_2} \left| \frac{\partial m^\dagger(z, \theta_n)}{\partial \theta_j} \cdot \frac{\partial m^\dagger(z, \theta_n)}{\partial \theta_i} - \frac{\partial m^\dagger(z, \theta_0)}{\partial \theta_j} \cdot \frac{\partial m^\dagger(z, \theta_0)}{\partial \theta_i} \right| w_z(\theta_0)
\]
\[ + \sum_{z=K+1}^{+\infty} \left| \frac{\partial m^\uparrow(z, \theta_n)}{\partial \theta_j} \cdot \frac{\partial m^\uparrow(z, \theta_n)}{\partial \theta_i} - \frac{\partial m^\uparrow(z, \theta_0)}{\partial \theta_j} \cdot \frac{\partial m^\uparrow(z, \theta_0)}{\partial \theta_l} \right| w_z(\theta_0) \]

\[ \leq \frac{\epsilon}{4} + \frac{\epsilon}{4}, \]

where we have bounded \( w_z(\theta_0) \) by 1 in the first sum and we have used Assumption \([A5]^{(a)}\) in the second sum. Because \( \lim_{n \to \infty} \mathbb{P}_i[B_{3n} | Z_n > 0] = 1 \) by the same argument as for \((36)\), we obtain \((38)\) for \( \ell = 2 \).

\[ \Box \]

**Acknowledgements**

Peter Braunsteins and Sophie Hautphenne would like to thank the Australian Research Council (ARC) for support through the Discovery Project DP200101281.

This manuscript was started while Carmen Minuesa was a visiting postdoctoral researcher in the School of Mathematics and Statistics at The University of Melbourne, and she is grateful for the hospitality and collaboration. She also acknowledges the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers for partially supporting her research visit at The University of Melbourne. Carmen Minuesa’s research is part of the R&D&I project PID2019-108211GB-I00, funded by MCIN/AEI/10.13039/501100011033/.

A Complement to the proof of Theorem 2

The following lemma establishes \((a-iii)\) in the proof of Theorem 2

**Lemma 4.** Under the conditions of Theorem 2 we have for \( \hat{w}_n(z) = \hat{w}_n^{(1)}(z) \) and \( \hat{w}_n(z) = \hat{w}_n^{(2)}(z) \)

\[ \lim_{k \to \infty} \left( \limsup_{n \to \infty} \mathbb{P}_i \left[ \sum_{z=k+1}^{+\infty} \frac{\partial m^\uparrow(z, \theta_0)}{\partial \theta_i} \hat{w}_n(z) \sqrt{n} (m^\uparrow(z, \theta_0) - \hat{m}_n(z)) \right| > \epsilon | Z_n > 0 \right] \right) = 0. \]

**Proof.** We establish the result when \( \hat{w}_n(z) = \hat{w}_n^{(2)}(z) \); the reasoning when \( \hat{w}_n(z) = \hat{w}_n^{(1)}(z) \) follows from similar (albeit simpler) arguments.

Fix \( j = 1, \ldots, d \) and denote

\[ V_{nz} = f(z) \hat{w}_n(z) \sqrt{n} (m^\uparrow(z, \theta_0) - \hat{m}_n(z)), \quad n, z \in \mathbb{N}, \]

where \( f(z) := \frac{\partial m^\uparrow(z, \theta_0)}{\partial \theta_j} \). We prove that for each \( \epsilon > 0 \),

\[ \lim_{k \to \infty} \left( \limsup_{n \to \infty} \mathbb{P}_i \left[ \sum_{z=k+1}^{+\infty} V_{nz} \right| > \epsilon | Z_n > 0 \right] \right) = 0. \]  \( (39) \)

We divide the proof of \( (39) \) in two steps. Recall the definitions of \( \{Z_{i}^{(n)}\} \) and \( \{Z_{i}^\uparrow\} \) from Section 2.2. For \( n, z \in \mathbb{N} \), let

\[ V_{nz}^\uparrow = f(z) q_n^\uparrow(z) \sqrt{n} (m^\uparrow(z, \theta_0) - \hat{m}_n(z)), \quad q_n^\uparrow(z) = \frac{z \sum_{i=0}^{n-1} \mathbbm{1}_{\{Z_{i+1}^\uparrow = z\}}}{\sum_{i=0}^{n-1} Z_{i}^\uparrow}, \quad \hat{m}_n^\uparrow(z) = \frac{\sum_{i=0}^{n-1} Z_{i+1}^\uparrow \mathbbm{1}_{\{Z_{i+1}^\uparrow = z\}}}{z \sum_{i=0}^{n-1} \mathbbm{1}_{\{Z_{i}^\uparrow\}}} \]

31
then, using Markov's inequality and the fact that

\[ V_{nz}^{(n)} = f(z)q_{zn}^{(n)}(z)\sqrt{n}(m^\uparrow(z, \theta_0) - \hat{m}_n^{(n)}(z)), \quad q_{zn}^{(n)}(z) = \frac{z \sum_{i=0}^{n-1} 1\{Z_i^{(n)} = z\}}{\sum_{i=0}^{n-1} Z_i^{(n)}} \quad \hat{m}_n^{(n)}(z) = \frac{\sum_{i=0}^{n-1} Z_{i+1}^{(n)} 1\{Z_i^{(n)} = z\}}{z \sum_{i=0}^{n-1} 1\{Z_i^{(n)} = z\}}. \]

**Step 1.** We show that the counterpart of (39) for the Q-process holds, that is,

\[ \forall \epsilon > 0, \quad \lim_{k \to \infty} \left( \limsup_{n \to \infty} \mathbb{P}_i \left[ \left| \sum_{n=k+1}^{\infty} V_{nz}^\uparrow \right| > \epsilon \right] \right) = 0, \quad (40) \]

**Step 2.** We prove that there exists a sequence of probability spaces (indexed by \( n \)) on which both \( \{Z_i^{(n)}\} \) and \( \{Z_i^\uparrow\} \) are defined such that

\[ \forall \epsilon > 0, \quad \lim_{k \to \infty} \left( \limsup_{n \to \infty} \mathbb{P}_i \left[ \left| \sum_{n=k+1}^{\infty} (V_{nz}^{(n)} - V_{nz}^\uparrow) \right| > \epsilon \right] \right) = 0, \quad (41) \]

We start with the proof of (40). First, observe that

\[
\sum_{z=k+1}^{\infty} V_{nz}^\uparrow = \sum_{z=k+1}^{\infty} f(z)q_{zn}^{(n)}(z)\sqrt{n}(m^\uparrow(z, \theta_0) - \hat{m}_n^{(n)}(z)) \\
= \sum_{z=k+1}^{\infty} f(z) z \sum_{i=0}^{n-1} 1\{Z_i^{(n)} = z\} \cdot \frac{\sqrt{n} \sum_{i=0}^{n-1} (zm^\uparrow(z, \theta_0) - Z_{i+1}^\uparrow) 1\{Z_i^{(n)} = z\}}{z \sum_{i=0}^{n-1} 1\{Z_i^{(n)} = z\}} \\
= \frac{1}{\sum_{i=0}^{n-1} Z_i^\uparrow} \sum_{z=k+1}^{\infty} f(z) \sqrt{n} \sum_{i=0}^{n-1} (zm^\uparrow(z, \theta_0) - Z_{i+1}^\uparrow) 1\{Z_i^{(n)} = z\},
\]

then, using Markov's inequality and the fact that \( \sum_{i=0}^{n-1} Z_i^\uparrow \geq n \) a.s., we have

\[
\mathbb{P}_i \left[ \left| \sum_{z=k+1}^{\infty} V_{nz}^\uparrow \right| > \epsilon \right] = \\
\geq \frac{1}{\epsilon^2} \mathbb{E}_i \left[ \left( \sum_{z=k+1}^{\infty} f(z) \sqrt{n} \sum_{i=0}^{n-1} (zm^\uparrow(z, \theta_0) - Z_{i+1}^\uparrow) 1\{Z_i^{(n)} = z\} \right)^2 \right] \\
\leq \frac{1}{\epsilon^2} \mathbb{E}_i \left[ \left( \sum_{z=k+1}^{\infty} f(z) \sqrt{n} \sum_{i=0}^{n-1} (zm^\uparrow(z, \theta_0) - Z_{i+1}^\uparrow) 1\{Z_i^{(n)} = z\} \right)^2 \right] \\
= \frac{1}{n \epsilon^2} \mathbb{E}_i \left[ \left( \sum_{z=k+1}^{\infty} f(z) \sum_{i=0}^{n-1} (zm^\uparrow(z, \theta_0) - Z_{i+1}^\uparrow) 1\{Z_i^{(n)} = z\} \right)^2 \right]. \quad (42)
\]
We now show that the expectation in (42) is finite, that is, letting \( a_l := \sum_{z=k+1}^{\infty} f(z)(zm^\dagger(z, \theta_0) - Z_{l+1}^\dagger) \mathbb{I}_{\{Z_l^\dagger = z\}} \) we show that \( \mathbb{E}_i \left[ (\sum_{l=0}^{n-1} a_l)^2 \right] < \infty \). We have

\[
\mathbb{E}_i \left[ \left( \sum_{l=0}^{n-1} a_l \right)^2 \right] = \sum_{l=0}^{n-1} \mathbb{E}_i [a_l^2] + 2 \sum_{j=0}^{n-1} \sum_{l=0}^{j-1} \mathbb{E}_i [a_l a_j].
\]

(43)

We start by showing that the second sum in (43) vanishes: for any fixed \( 0 \leq j \leq n - 1 \) and \( 0 \leq l \leq j - 1 \),

\[
\mathbb{E}_i [a_l a_j] = \sum_{z_1=k+1}^{\infty} \sum_{z_2=k+1}^{\infty} \mathbb{E}[a_l a_j | Z_l^\dagger = z_1, Z_j^\dagger = z_2] \mathbb{P}_i[Z_l^\dagger = z_1, Z_j^\dagger = z_2]
\]

\[
= \sum_{z_1=k+1}^{\infty} \sum_{z_2=k+1}^{\infty} f(z_1) f(z_2) \mathbb{E}[(zm^\dagger(z_1, \theta_0) - Z_{l+1}^\dagger)(zm^\dagger(z_2, \theta_0) - Z_{j+1}^\dagger) | Z_l^\dagger = z_1, Z_j^\dagger = z_2]
\]

\[
\times \mathbb{P}_i[Z_l^\dagger = z_1, Z_j^\dagger = z_2]
\]

\[= 0, \]

where the last equality follows from the fact that \( \mathbb{E}[(zm^\dagger(z, \theta_0) - Z_{l+1}^\dagger) | Z_l^\dagger = z] = 0 \) for any \( z > 0 \), and the conditional independence of \( Z_l^\dagger \) and \( Z_j^\dagger \) given \( Z_l^\dagger = z_1 \) and \( Z_j^\dagger = z_2 \).

Now we treat the first sum in (43):

\[
\mathbb{E}_i [a_l^2] = \sum_{x=k+1}^{\infty} \mathbb{E}[a_l^2 | Z_l^\dagger = x] \mathbb{P}_i[Z_l^\dagger = x]
\]

\[
= \sum_{x=k+1}^{\infty} f(x)^2 \mathbb{E}[(xm^\dagger(x, \theta_0) - Z_{l+1}^\dagger)^2 | Z_l^\dagger = x] \mathbb{P}_i[Z_l^\dagger = x]
\]

\[
\leq M_l^\dagger(\theta_0)^2 \sum_{x=k+1}^{\infty} \text{Var}[Z_{l+1}^\dagger | Z_l^\dagger = x] \mathbb{P}_i[Z_l^\dagger = x]
\]

where in the last inequality we used Assumption (A5)(a) (and the fact that \( \mathbb{E}[(xm^\dagger(x, \theta_0) - Z_{l+1}^\dagger) | Z_l^\dagger = x] = 0 \)). We then have from (42) and (43)

\[
\mathbb{P}_i \left[ \left| \sum_{z=k+1}^{\infty} V_{nz}^\dagger \right| > \epsilon \right] \leq \frac{1}{n \epsilon^2} \sum_{l=0}^{n-1} \mathbb{E}_i [a_l^2]
\]

\[
\leq \frac{M_l^\dagger(\theta_0)^2}{n \epsilon^2} \sum_{z=k+1}^{\infty} \sum_{l=0}^{n-1} \text{Var}[Z_l^\dagger | Z_l^\dagger = z] \mathbb{P}_i[Z_l^\dagger = z]
\]

\[
= \frac{M_l^\dagger(\theta_0)^2}{n \epsilon^2} \sum_{z=k+1}^{\infty} z^2 \sigma_{z}^2(\theta_0)(Q^\dagger)_{iz}(\theta_0)
\]

\[
= \frac{M_l^\dagger(\theta_0)^2}{n \epsilon^2 v_i(\theta_0)} \sum_{z=k+1}^{\infty} z^2 \sigma_{z}^2(\theta_0) v_z(\theta_0) \sum_{l=0}^{n-1} \frac{(Q)_{iz}(\theta_0)}{\rho^l}
\]

33
\[
\frac{M_t^2(\theta_0)}{\sigma^2v_i(\theta_0)} \sum_{z=k+1}^\infty z^2 \sigma^2(z, \theta_0)v_z(\theta_0) \sum_{l=0}^{n-1} \frac{e_i^\top Q(\theta_0) e_z}{\rho^l} = \frac{M_t^2(\theta_0)}{\sigma^2v_i(\theta_0)} \sum_{z=k+1}^\infty z^2 \sigma^2(z, \theta_0)v_z(\theta_0) \sum_{l=0}^{n-1} \left( v_i(\theta_0)u_z(\theta_0) + \frac{e_i^\top S(\theta_0) e_z}{\rho^l} \right) \\
= \frac{M_t^2(\theta_0)}{\sigma^2v_i(\theta_0)} \sum_{z=k+1}^\infty z^2 \sigma^2(z, \theta_0)v_z(\theta_0) \sum_{l=0}^{n-1} \left( v_i(\theta_0)u_z(\theta_0) + \frac{e_i^\top S(\theta_0) e_z}{\rho^l} \right) \\
+ \frac{M_t^2(\theta_0)}{\sigma^2v_i(\theta_0)} \sum_{z=k+1}^\infty z^2 \sigma^2(z, \theta_0)v_z(\theta_0) \sum_{l=0}^{n-1} \left( v_i(\theta_0)u_z(\theta_0) + \frac{e_i^\top S(\theta_0) e_z}{\rho^l} \right) \\
= \frac{M_t(\theta_0)}{\sigma^2v_i(\theta_0)} \sum_{z=k+1}^\infty z^2 \sigma^2(z, \theta_0)v_z(\theta_0) \sum_{l=0}^{n-1} \left( v_i(\theta_0)u_z(\theta_0) + \frac{e_i^\top S(\theta_0) e_z}{\rho^l} \right) \\
+ \frac{M_t^2(\theta_0)}{\sigma^2v_i(\theta_0)} \sum_{z=k+1}^\infty z^2 \sigma^2(z, \theta_0)v_z(\theta_0) \sum_{l=0}^{n-1} \left( v_i(\theta_0)u_z(\theta_0) + \frac{e_i^\top S(\theta_0) e_z}{\rho^l} \right). \\
\]

Now, we first show that \( \sum_{z=1}^\infty z^2 \sigma^2(z, \theta_0)v_z(\theta_0)u_z(\theta_0) < \infty \) so that the term \((44)\) vanishes as \( k \to \infty \). In what follows, we use the function \( t(\cdot) \) and the norms \( \| \cdot \|_{\infty,t} \) and \( \| \cdot \|_{1,t} \) defined in [2] Section 5.3.1. We have

\[
\sum_{z=1}^\infty z^2 u_z(\theta_0)v_z(\theta_0) \sigma^2(z, \theta_0) \leq \sum_{z=1}^\infty z^2 u_z(\theta_0)v_z(\theta_0) \sum_{k=1}^\infty k^2 Q_z^2(\theta_0) \\
= \sum_{z=1}^\infty u_z(\theta_0)v_z(\theta_0) \sum_{k=1}^\infty k^2 Q_z^2(\theta_0) \\
= \frac{1}{\rho(\theta_0)} \sum_{z=1}^\infty u_z(\theta_0) \sum_{k=1}^\infty k^{l+1} Q_z^2(\theta_0) \\
\leq \frac{\| \mathbf{v}(\theta_0) \|_{\infty,t}}{\rho(\theta_0)} \sum_{z=1}^\infty u_z(\theta_0) \sum_{k=1}^\infty k^{l+1} Q_z^2(\theta_0) \\
= \frac{\| \mathbf{v}(\theta_0) \|_{\infty,t}}{\rho(\theta_0)} \sum_{z=1}^\infty u_z(\theta_0) E[Z_1^{l+3} | Z_0 = z] \\
\leq C \frac{\| \mathbf{v}(\theta_0) \|_{\infty,t} \| \mathbf{u}(\theta_0) \|_{1,t} }{\rho(\theta_0)} < \infty,
\]

for some constant \( C > 0 \), where we applied [2] Lemma 13] with the functions \( t(x) = x^{l+1} \) and \( t^*(x) = x^{l+3} \), Minkowski’s inequality, and Assumption [C3].
Finally, to obtain (40) it remains to treat (45), i.e., to prove that
\[
\lim_{k \to \infty} \left( \limsup_{n \to \infty} \left( \frac{M_1^1(\theta_0)^2}{n \epsilon^2 v_1(\theta_0)} \sum_{z=k+1}^\infty z^2 \sigma^{z^2}(z, \theta_0) v_z(\theta_0) \sum_{l=0}^{n-1} \frac{e_i^T S(\theta_0) e_z}{\rho^l} \right) \right) = 0. \quad (46)
\]

To show this, we make use of [2, Inequality (51)], that is, the fact that there exist \( K < \infty \) and \( \delta > 0 \) such that \( |e_i^T S(\theta_0) e_z| \leq \frac{t(i)}{t(z)} K (1-\delta)^l \rho^l \). We have
\[
\frac{1}{n \epsilon^2 v_1(\theta_0)} \sum_{z=k+1}^\infty z^2 \sigma^{z^2}(z, \theta_0) v_z(\theta_0) \sum_{l=0}^{n-1} \frac{|e_i^T S(\theta_0) e_z|}{\rho^l} \leq \frac{t(i)K}{n \epsilon^2 v_1(\theta_0)} \sum_{z=k+1}^\infty z^2 \sigma^{z^2}(z, \theta_0) v_z(\theta_0) \sum_{l=0}^{n-1} \frac{(1-\delta)^l \rho^l}{\rho^l t(z)} \leq \frac{t(i)K}{n \epsilon^2 v_1(\theta_0)} \frac{1-(1-\delta)^n}{\delta} \sum_{z=k+1}^\infty z^2 \sigma^{z^2}(z, \theta_0) v_z(\theta_0) \frac{t(z)}{t(z)} \leq \frac{t(i)K}{n \epsilon^2 v_1(\theta_0)} \frac{1-(1-\delta)^n}{\delta} \sum_{z=k+1}^\infty \sum_{l=1}^\infty l^2 Q_{zl}(\theta_0) v_z(\theta_0) \frac{1}{t(z)}, \quad (47)
\]
where in the last step we used the definition of \( \sigma^{z^2}(z, \theta_0) \) and the fact that \( \text{Var}[X] \leq \mathbb{E}[X^2] \).

Next we use [2, Lemma 13] which implies that \( v_i(\theta_0) < C l^{\nu^*} \) for all \( l \geq 1 \) for some positive constants \( C \) and \( \nu^* \) and which justifies the choice of \( t(z) = z^{\nu^*+m'} \) for some positive constant \( m' \):
\[
(47) = \frac{t(i)K \left[ 1-(1-\delta)^n \right]}{n \epsilon^2 v_1(\theta_0) \delta \rho} \sum_{z=k+1}^\infty \sum_{l=1}^\infty l^2 Q_{zl}(\theta_0) \frac{v_l(\theta_0) v_z(\theta_0)}{t(z)} = \frac{t(i)K \left[ 1-(1-\delta)^n \right]}{n \epsilon^2 v_1(\theta_0) \delta \rho} \sum_{z=k+1}^\infty \sum_{l=1}^\infty l^2 Q_{zl}(\theta_0) \frac{1}{z^{\nu^*+m'}} \leq \frac{t(i)K \left[ 1-(1-\delta)^n \right] C}{n \epsilon^2 v_1(\theta_0) \delta \rho} \sum_{z=k+1}^\infty \sum_{l=1}^\infty l^{2+\nu^*} Q_{zl}(\theta_0) \frac{1}{z^{\nu^*+m'}}. \quad (48)
\]

Now we use Minkowski’s inequality and Assumption (C3) to write
\[
\sum_{l=1}^\infty l^{2+\nu^*} Q_{zl}(\theta_0) = \mathbb{E}[Z_1^{2+\nu^*} \mid Z_0 = z] \leq z^{2+\nu^*} \mathbb{E}[\xi^{2+\nu^*}] \leq z^{2+\nu^*} C''
\]
for some constant \( C'' \). By choosing \( m' = 4 \) we obtain
\[
(48) \leq \frac{t(i)K \left[ 1-(1-\delta)^n \right] C C''}{n \epsilon^2 v_1(\theta_0) \delta \rho} \sum_{z=k+1}^\infty z^{2+\nu^*} \frac{z^{\nu^*+4}}{z^{\nu^*+4}} < \infty,
\]

35
where the last sum is finite since it is bounded by \( \sum_{z=1}^{\infty} \frac{1}{z^2} < \infty \).

We now move on to Step 2. According to [2, Theorem 3.1(i)], there exists a sequence of probability spaces (indexed by \( n \)) such that for any \( K \),

\[
\lim_{n \to \infty} \inf \mathbb{P}_t(Z_j^{(n)} = Z_j^\uparrow, \forall j = 1, \ldots, n - K) = 1 - \frac{\rho^{-K}}{2} \sum_{j=1}^{\infty} u_j|e_j^\uparrow S^K 1|.
\]  

(49)

Next, using [2, Inequality (52)] which states that there exists \( L < \infty \) and \( \varepsilon > 0 \) such that, for any function \( t \) identified in [2, Lemma 13], \( |e_j^\uparrow S^K 1| \leq t(j) L(1 - \varepsilon)^K \rho^K \), we get

\[
(49) \geq 1 - \frac{1}{2} L(1 - \varepsilon)^K \sum_{j=1}^{\infty} u_j t(j),
\]

where \( \sum_{j=1}^{\infty} u_j t(j) < \infty \) because \( u \in l^1 := \{ x \in \mathbb{R}^N : ||x||_{1,1} < \infty \} \) (see [2, Section 5.3.1]). Therefore, for any \( \eta > 0 \) there exists \( K(\eta) \in \mathbb{N} \) such that

\[
\lim_{n \to \infty} \inf \mathbb{P}_t(Z_j^{(n)} = Z_j^\uparrow, \forall j = 1, \ldots, n - K) \geq 1 - \eta/3.
\]

In addition, for any \( K(\eta) \in \mathbb{N} \), both \( \sum_{j=n-K(\eta)}^{n} Z_j^{(n)} \) and \( \sum_{j=n-K(\eta)}^{n} Z_j^\uparrow \) converge to non-defective random variables as \( n \to \infty \). Indeed, for \( \sum_{j=n-K(\eta)}^{n} Z_j^{(n)} \) this follows from [Gosselin, Theorem 3.1(b)], and for \( \sum_{j=n-K(\eta)}^{n} Z_j^\uparrow \) this follows from the ergodicity of \( \{Z_n^\uparrow\} \). Consequently for any \( \eta > 0 \) and \( K(\eta) \in \mathbb{N} \) there exists \( M(K, \eta) \) such that

\[
\limsup_{n \to \infty} \mathbb{P}\left( \sum_{j=n-K(\eta)}^{n} Z_j^{(n)} > M(K, \eta) \right) \leq \eta/3 \quad \text{and} \quad \limsup_{n \to \infty} \mathbb{P}\left( \sum_{j=n-K(\eta)}^{n} Z_j^\uparrow > M(K, \eta) \right) \leq \eta/3.
\]

We define the event

\[
E_n^* := \{ Z_j^{(n)} = Z_j^\uparrow, \forall j = 1, \ldots, n - K(\eta) \}\]

\[
\cap \left\{ \sum_{j=n-K(\eta)}^{n} Z_j^{(n)} \leq M(K, \eta) \right\} \cap \left\{ \sum_{j=n-K(\eta)}^{n} Z_j^\uparrow \leq M(K, \eta) \right\}.
\]

For any \( \eta > 0 \) there then exists \( K(\eta) \in \mathbb{N} \) and \( M(K, \eta) \) such that, on the sequence of probability spaces described in [2, Theorem 3.1], we have

\[
\lim_{n \to \infty} \inf \mathbb{P}(E_n^*) \geq \lim_{n \to \infty} \left[ 1 - \mathbb{P}\left( \exists j \in \{1, \ldots, n - K(\eta)\} : Z_j^{(n)} \neq Z_j^\uparrow \right) - \mathbb{P}\left( \sum_{j=n-K(\eta)}^{n} Z_j^{(n)} > M(K, \eta) \right) - \mathbb{P}\left( \sum_{j=n-K(\eta)}^{n} Z_j^\uparrow > M(K, \eta) \right) \right] \geq 1 - \eta.
\]
It therefore suffices to show that, on the event $E_n^*$, we have
\[
\left| \sum_{z=k}^{\infty} (V_{nz}^{(n)} - V_{nz}^\uparrow) \right| \to 0, \quad \text{as } n \to \infty, \tag{50}
\]
for any $k, K(\eta), M(K, \eta) \in \mathbb{N}$, because in this case we have
\[
\limsup_{n \to \infty} \mathbb{P}_i \left[ \sum_{z=k+1}^{\infty} (V_{nz}^{(n)} - V_{nz}^\uparrow) > \varepsilon \right] \leq 1 - \liminf_{n \to \infty} \mathbb{P}(E_n^*) \leq \eta,
\]
where, since $\eta$ was chosen arbitrarily, we can let $\eta \downarrow 0$.

We now establish (50). We have
\[
\sum_{z=k}^{\infty} (V_{nz}^{(n)} - V_{nz}^\uparrow) = \sqrt{n} \sum_{z=k}^{\infty} f(z) \left[ \hat{q}_n^{(n)}(z)(m^\uparrow(z, \theta_0) - \hat{m}_n^{(n)}(z)) - q_n^*(z)(m^\uparrow(z, \theta_0) - \hat{m}_n^*(z)) \right]
= \sqrt{n} \sum_{z=k}^{\infty} f(z) \left[ \{ (q_n^*(z) - q_n^{(n)}(z))m^\uparrow(z, \theta_0) \} + \{ \hat{m}_n^{(n)}(z)q_n^{(n)} - \hat{m}_n^*(z)q_n^*(z) \} \right].
\]

We bound the sums resulting from these two terms separately. We start with the second term. Using the assumption that we are on the event $E_n^*$ in steps 4 to 6, we obtain
\[
\begin{align*}
\left| \sum_{z=k}^{\infty} f(z) \left[ \hat{m}_n^{(n)}(z)q_n^{(n)} - \hat{m}_n^*(z)q_n^*(z) \right] \right| & = \left| \sum_{z=k}^{\infty} f(z) \left[ \frac{\sum_{i=0}^{n-1} Z_i^{(n)} \mathbb{1}_{\{Z_i^{(n)} = z\}}}{\sum_{i=0}^{n-1} Z_i^{(n)}} - \frac{\sum_{i=0}^{n} Z_i^\uparrow \mathbb{1}_{\{Z_i^\uparrow = z\}}}{\sum_{i=0}^{n} Z_i^\uparrow} \right] \right| \\
& = \left| \sum_{z=k}^{\infty} f(z) \left[ \left( \sum_{i=0}^{n-1} Z_i^{(n)} \mathbb{1}_{\{Z_i^{(n)} = z\}} \right) \left( \sum_{i=0}^{n} Z_i^\uparrow \right) - \left( \sum_{i=0}^{n} Z_i^\uparrow \mathbb{1}_{\{Z_i^\uparrow = z\}} \right) \left( \sum_{i=0}^{n-1} Z_i^{(n)} \right) \right] \right| \\
& = \left| \frac{1}{\left( \sum_{i=0}^{n-1} Z_i^{(n)} \right) \left( \sum_{i=0}^{n} Z_i^\uparrow \right)} \sum_{z=k}^{\infty} f(z) \left[ \left( \sum_{i=0}^{n-1} Z_i^{(n)} \mathbb{1}_{\{Z_i^{(n)} = z\}} \right) - \left( \sum_{i=0}^{n} Z_i^\uparrow \mathbb{1}_{\{Z_i^\uparrow = z\}} \right) \right] \right| \\
& \leq M^*(\theta_0) \left\{ \left| \sum_{i=0}^{n-1} (Z_i^\uparrow - Z_i^{(n)}) \right| \frac{1}{\sum_{i=0}^{n} Z_i^\uparrow} \left( \sum_{i=0}^{n} Z_i^\uparrow \right) \right\} + \left( \sum_{z=k}^{\infty} \frac{\sum_{i=0}^{n} Z_i^\uparrow \mathbb{1}_{\{Z_i^\uparrow = z\}}}{\sum_{i=0}^{n} Z_i^\uparrow} \right) \left| \sum_{i=0}^{n-1} Z_i^{(n)} \right| \\
& \leq 3M^*(\theta_0)M(K, \eta)/n,
\end{align*}
\]

37
where in the penultimate step we have applied the triangular inequality, and Assumption \((A5)(a)\), and in the final step, we have applied the inequalities \(\sum_{i=0}^{n-1} Z_i^{(n)} \geq n\) and \(\sum_{i=0}^{n-1} Z_i^\uparrow \geq n\) and the fact that we are on \(E_n^*\). Applying similar arguments (which are therefore omitted) we obtain the same bound for the first term. We have thus established the result.

References

[1] P. Billingsley. *Probability and Measure*. John Wiley and Sons, Inc., 1995.

[2] P. Braunsteins, S. Hautphenne, and C. Minuesa. Parameter estimation in branching processes with almost sure extinction. *Bernoulli*, 28(1):33–63, 2022.

[3] D. Butler and D. Merton. *The Black Robin: Saving the World’s Most Endangered Bird*. Oxford University Press, 1992.

[4] E. P. Elliston. The black robin: saving the world’s most endangered bird. *BioScience*, 44(7):499–501, 1994.

[5] M. González and I. del Puerto. Weighted conditional least squares estimation in controlled multitype branching processes. In M. González, I. del Puerto, R. Martínez, M. Molina, M. Mota, and A. Ramos, editors, *Workshop on Branching Processes and Their Applications*, volume 197 of *Lecture Notes in Statistics-Proceedings*, pages 147–155. Springer, 2010.

[6] M. González, C. Minuesa, and I. del Puerto. Approximate Bayesian computation approach on the maximal offspring and parameters in controlled branching processes. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas*, 116:147, 2022.

[7] F. Gosselin. Asymptotic behaviour of absorbing Markov chains conditional on nonabsorption for applications in conservation biology. *The Annals of Applied Probability*, 11(1):261–284, 2001.

[8] P. Guttorp. *Statistical inference for branching processes*, volume 122. Wiley-Interscience, 1991.

[9] S. Hautphenne and B. Patch. Birth-and-death processes in Python: The BirDePy package. *arXiv preprint arXiv:2110.05067*, 2021.

[10] P. Jagers. Stabilities and instabilities in population dynamics. *Journal of Applied Probability*, 29:770–780, 1992.

[11] P. Jagers and S. Zuyev. Populations in environments with a soft carrying capacity are eventually extinct. *Journal of Mathematical Biology*, 81(3):845–851, 2020.

[12] E. S. Kennedy, C. E. Grueber, R. P. Duncan, and I. G. Jamieson. Severe inbreeding depression and no evidence of purging in an extremely inbred wild species—the Chatham Island black robin. *Evolution*, 68(4):987–995, 2014.
[13] L. A. Klimko and P. I. Nelson. On conditional least squares estimation for stochastic processes. *The Annals of Statistics*, 6(3):629–642, 1978.

[14] N. Lalam and C. Jacob. Estimation of the offspring mean in a supercritical or near-critical size-dependent branching process. *Advances in Applied Probability*, 36(2):582–601, 2004.

[15] N. Lalam, C. Jacob, and P. Jagers. Modelling the PCR amplification process by a size-dependent branching process and estimation of the efficiency. *Advances in Applied Probability*, 36(2):602–615, 2004.

[16] R. Lande, S. Engen, and B. E. Saether. *Stochastic population dynamics in ecology and conservation*. Oxford University Press on Demand, 2003.

[17] M. Massaro, A. Chick, E. S. Kennedy, and R. Whitsed. Post-reintroduction distribution and habitat preferences of a spatially limited island bird species. *Animal Conservation*, 21(1):54–64, 2018.

[18] M. Massaro, M. Stanbury, and J. V. Briskie. Nest site selection by the endangered black robin increases vulnerability to predation by an invasive bird. *Animal Conservation*, 16(4):404–411, 2013.

[19] A. G. Pakes. Non-parametric estimation in the Galton-Watson process. *Mathematical Biosciences*, 26(1):1–18, 1975.