University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Hautphenne, S;Massaro, M;Taylor, P

Title:
How old is this bird? The age distribution under some phase sampling schemes

Date:
2017-12-01

Citation:
Hautphenne, S., Massaro, M. & Taylor, P. (2017). How old is this bird? The age distribution under some phase sampling schemes. Journal of Mathematical Biology, 75 (6-7), pp.1319-1347. https://doi.org/10.1007/s00285-017-1121-x.

Persistent Link:
https://hdl.handle.net/11343/249510

Noname manuscript No.
(will be inserted by the editor)

How old is this bird?

The age distribution under some phase sampling schemes

Sophie Hautphenne · Melanie Massaro · Peter Taylor

Received: date / Accepted: date

Abstract In this paper, we use a finite-state continuous-time Markov chain with one absorbing state to model an individual's lifetime. Under this model, the time of death follows a phase-type distribution, and the transient states of the Markov chain are known as *phases*. We then attempt to provide an answer to the simple question "*What is the conditional age distribution of the individual, given its current phase*"? We show that the answer depends on how we interpret the question, and in particular, on the phase observation scheme under consideration. We then apply our results to the computation of the age pyramid for the endangered Chatham Island black robin *Petroica traversi* during the monitoring period 2007-2014.

Keywords Phase-type distribution · Transient Markov chain · Age distribution · *Petroica traversi*

Mathematics Subject Classification (2000) 60J27 · 92B05 · 92D25

1 Introduction

A random variable has a phase-type (PH) distribution if it corresponds to the time until absorption of a transient Markov chain with one absorbing

S. Hautphenne
School of Mathematics and Statistics, The University of Melbourne, Melbourne, VIC 3010, Australia
Tel.: +61 3 8344 9073
Fax: +61 3 8344 4599
E-mail: sophiemh@unimelb.edu.au

M. Massaro
Institute of Land, Water and Society, and the School of Environmental Sciences, Charles Sturt University, Albury, NSW 2640, Australia

P. Taylor
School of Mathematics and Statistics, The University of Melbourne, Melbourne, VIC 3010, Australia

state. PH distributions, introduced in the early 1980's by Neuts [12, Chapter 2], form a class of distributions with considerable modelling versatility, which results from attractive probabilistic properties. The set of PH distributions is closed under convolutions and finite mixtures, and is dense in the class of all distributions with non-negative support. PH distributions have therefore been extensively used in practice, in particular for modelling lifetime distributions, see for instance [1], [5], and [9].

The question addressed in this paper arose initially when modelling the lifetime reproductive success of the black robin *Petroica traversi*, which is an endangered songbird species endemic to the Chatham Islands, an isolated archipelago located 800 kilometers East of New Zealand. By 1980, the population of black robins had declined to five birds, including only a single successful breeding pair [2]. Through intensive conservation efforts in 1980-1989 by the New Zealand Wildlife Service (now the Department of Conservation), the population recovered to 93 birds by spring 1990 [8]. Over the next decade (1990-1998), the population was closely monitored, but without human intervention. Nevertheless the population continued to grow rapidly to 197 adults by 1998, but after this period, the population growth slowed considerably and it only reached 239 adults in 2011 [11] and 298 in 2014.

The black robin population is modeled in [7] using a special class of branching processes called the Markovian binary tree (MBT), in which an underlying transient Markov chain controls the reproduction and death events of each individual in the population. A direct consequence of this model is that each bird lives for a random time which has a PH distribution, in which it progresses through states (also called phases) of a continuous-time Markov chain and dies when the chain moves to an absorbing state. In this application, the phases do not have any particular physical interpretation, their role is to increase the accuracy and realism of the MBT model, as opposed to the simplest linear birth-and-death model. In other real-world applications, the phases may have a physical meaning, such as in [9] where they model physiological ages, which can be interpreted as relative health indices, as opposed to chronological age.

MBTs have proved to be powerful stochastic models in population biology and demography [6]. Having fitted an MBT to real data, we can calculate properties of the population, such as the probability that it will become extinct in some time interval $[0, t]$, and the distribution of the population size at time t. In particular, the model allows us to compute the expected number of birds in phase j at time t, and the asymptotic frequency of phase j in the population. However, the latter two quantities may not have significance for biologists, who are likely to be interested in age-specific, rather than phase-specific, properties of individuals. We therefore need to be able to translate information about the phase distribution into information about the age distribution. Once we have an answer to the question "*what is the age distribution of a bird, given its phase?*", we can write expression for age-dependent quantities in terms of phase-dependent quantities. For example, Equation (35) in Section 7.2 illustrates how to translate the asymptotic phase frequency into the asymptotic age-frequency.

Note that the reverse exercise of translating information about age into information about phase is much easier since the distribution of the phase at any given age is well known. The main difference is that age is deterministic, while phase is random and an individual stays in a given phase for an exponentially distributed amount of time.

In fact, defining the event that an individual is in phase j is already not trivial. It is necessary to describe in more detail how the individual is sampled. We suggest three sampling schemes according to which an observer looks at the phase of an individual:

- first, we assume that individuals are born according to a Poisson process which started infinitely far in the past, and one observes the phase of a randomly-selected individual that is still alive at time 0.

In the second and third schemes, we include the observation process in the lifetime history of each bird, which means that we do not need to make any assumption about the birth process:

- in the second scheme, observations of the phase of the bird occur according to a Poisson process, in which case we allow a single observation or multiple observations;
- in the third scheme, a single observation occurs at a uniformly distributed random time within some time window.

For each observation scheme, we compute the conditional age distribution given the observed phase, as well as related quantities. We show that the age distribution conditional on a single rare Poisson phase observation coincides with the age distribution conditional on a single rare uniform phase observation. Moreover, this age-distribution also corresponds to the age distribution of a randomly selected bird in a given phase at time 0 in the process where individuals are born according to a Poisson process.

The questions addressed in this paper, and their proposed answers, are not restricted to the context of PH distributions and ageing processes; they have a wider interest in Markov chain theory. Indeed, if Q is the generator of a continuous-time Markov chain $\{X(t) : t \geq 0\}$, then it is well known that $P[X(t) = j|X(0) = i] = (e^{Qt})_{ij}$, but the conditional distribution of the time t elapsed since the start of the Markov chain, given that the chain is observed in phase j is much less explored. As we mentioned above, the nature of this observation event needs to be described carefully.

The paper is organised as follows. In the next section, we provide some background on PH distributions used to model the ageing process of individuals. In Section 3, we compute the conditional age distribution at time 0 in the Poisson birth process, given the observed phase. In section 4, we consider the Poisson phase observation scheme and provide the conditional age distribution with a single observation or multiple observations. In Section 5, we consider a uniform observation scheme, and in Section 6 we discuss the rare observation limit of the results obtained in Sections 4 and 5. Finally, in Section 7, we illustrate our results on a toy example first, and then on the computation of the age pyramid for the black robin population.

2 The phase-type lifetime distribution

We assume that the lifetime of an individual is a random variable L which follows a phase-type $\mathrm{PH}(\boldsymbol{\alpha}, \mathbf{Q})$ distribution with m transient phases $\{1, 2, \ldots, m\}$ and the absorbing phase 0. This PH distribution is parameterised by an $1 \times m$ vector $\boldsymbol{\alpha}$ which gives us the initial distribution of the underlying Markov chain, and an $m \times m$ matrix $\mathbf{Q}$ containing the transition rates between the transient phases. So the assumption is that the lifetime of an individual progresses through phases (which may or may not correspond to some physically-observable characteristics) according to a realisation of the Markov chain, and the individual dies when the chain moves to the absorbing phase 0.

The $\mathrm{PH}(\boldsymbol{\alpha}, \mathbf{Q})$ distributed random variable L has a density and a distribution function respectively given by

$$\begin{aligned} f_L(x) &= \boldsymbol{\alpha} e^{\mathbf{Q}x} \boldsymbol{q}_0 \\ F_L(x) = \mathrm{P}[L \leq x] &= 1 - \boldsymbol{\alpha} e^{\mathbf{Q}x} \mathbf{1}, \end{aligned} \tag{1}$$

where $\boldsymbol{q}_0 = (-\mathbf{Q})\mathbf{1}$ is the absorption rate vector, and $\mathbf{1}$ is a column vector of ones. Let $\varphi(x)$ denote the phase of the individual at age x, and let j be any transient phase. Another basic result on PH random variables tells us that the probability that an individual is in phase j when its age is x is

$$\mathrm{P}[\text{phase} = j \,|\, \text{age} = x] = \mathrm{P}[\varphi(x) = j] = (\boldsymbol{\alpha} e^{\mathbf{Q}x})_j = \boldsymbol{\alpha} e^{\mathbf{Q}x} \boldsymbol{e}_j,$$

where $\boldsymbol{e}_j$ is the jth unit (column) vector. Our question is the reverse of this: "*If we observe an individual in phase j, what can we say about its age?*" Bayes' Theorem gives us

$$\mathrm{P}[\text{age} \leq x \,|\, \text{phase} = j] = \frac{\mathrm{P}[\text{phase} = j \,|\, \text{age} \leq x]\, \mathrm{P}[\text{age} \leq x]}{\mathrm{P}[\text{phase} = j]}.$$

The problem is that we do not yet have anything in the model to make sense of $\mathrm{P}[\text{age} \leq x]$ and $\mathrm{P}[\text{phase} = j]$. Our aim in this paper is to find satisfying answers to the above question.

For further use, we denote the age of the individual at the time of observation as A_o and the observed phase as φ_o. We are therefore interested in computing

$$\mathrm{P}_s[A_o \leq x \,|\, \varphi_o = j], \tag{2}$$

where the subscript in P_s corresponds to the observation scheme that we shall consider: $s = pb$ for the *Poisson birth* process (Section 3), $s = po$ for the *Poisson observation* scheme (Section 4), and $s = uo$ for the *uniform observation* scheme (Section 5).

3 A Poisson birth process

One assumption that we might make is that individuals have been born at the epochs of a Poisson process with parameter β over the time interval $(-\infty, 0)$, and that we observe the phase of a randomly-selected individual alive at time 0.

Lemma 31 *For any $x \geq 0$,*

$$P_{pb}[A_o \leq x \,|\, \varphi_o = j] = 1 - \frac{\boldsymbol{\alpha} e^{\mathbf{Q}x}(-\mathbf{Q})^{-1}\boldsymbol{e}_j}{\boldsymbol{\alpha}(-\mathbf{Q})^{-1}\boldsymbol{e}_j}. \tag{3}$$

Proof. For any $T > 0$,

- the number N_T of individuals born in the time interval $[-T, 0)$ has a Poisson distribution with parameter βT,
- conditional on $N_T = n$, the birthtimes of the n individuals are uniformly and independently distributed on the interval $[-T, 0)$,
- an individual born at time $u \in [-T, 0)$ will be alive at time 0 with probability $\boldsymbol{\alpha} e^{\mathbf{Q}(-u)}\mathbf{1}$ by (1), independently of the event that any other individual is alive at time 0.

So, an individual born in the interval $[-T, 0)$ will be alive at time 0 with probability $(1/T)\int_{-T}^{0} \boldsymbol{\alpha} e^{\mathbf{Q}(-u)}\mathbf{1}\, du$. It will be alive and in phase j at time 0 with probability $(1/T)\int_{-T}^{0} \boldsymbol{\alpha} e^{\mathbf{Q}(-u)}\boldsymbol{e}_j\, du$, and it will be alive, in phase j and older than x at time 0 with probability $(1/T)\int_{-T}^{-x} \boldsymbol{\alpha} e^{\mathbf{Q}(-u)}\boldsymbol{e}_j\, du$, again independently of what happens to any other individual born in $[-T, 0)$. So the probability that an individual in phase j at time 0 is older than x is

$$\frac{\int_{-T}^{-x} \boldsymbol{\alpha} e^{\mathbf{Q}(-u)}\boldsymbol{e}_j\, du}{\int_{-T}^{0} \boldsymbol{\alpha} e^{\mathbf{Q}(-u)}\boldsymbol{e}_j\, du}.$$

The fact that any PH distribution has a finite mean allows us to let $T \to \infty$ and, changing the variable of integration, we arrive at the conclusion that the probability that a randomly-selected individual in phase j at time 0 is older than x is

$$\frac{\int_{x}^{\infty} \boldsymbol{\alpha} e^{\mathbf{Q}u}\boldsymbol{e}_j\, du}{\int_{0}^{\infty} \boldsymbol{\alpha} e^{\mathbf{Q}u}\boldsymbol{e}_j\, du} = \frac{\boldsymbol{\alpha} e^{\mathbf{Q}x}(-\mathbf{Q})^{-1}\boldsymbol{e}_j}{\boldsymbol{\alpha}(-\mathbf{Q})^{-1}\boldsymbol{e}_j},$$

which completes the proof. □

Note that this model corresponds to an $M/PH/\infty$ queue, and (3) gives the distribution of the age of a randomly selected individual in steady state. Also observe that $\mathrm{P}_{pb}[A_o \leq x \,|\, \varphi_o = j]$ does not depend on the rate β of the Poisson birth process.

The above analysis provides a neat formula for the age distribution conditional on the observed phase. However the birth process in an MBT is not Poisson and, more generally, there is no reason to believe that a Poisson process is a good model for births. In the next three sections, we shall follow an alternative approach: without making any assumption on the birth process, we look at just a single individual and proceed by explicitly putting the observation process into the model.

4 The Poisson observation scheme

In this section, we assume that, following its birth, the phase of an individual is observed according to a Poisson process with rate γ. We first compute the age distribution given the phase at the first observation time. Then we generalise our results to the age distribution at the time of the last observation, given that the observer records the phases at $k \geq 2$ successive time events of the Poisson process.

4.1 Single observation

By the properties of Poisson processes, the rate at which an individual is observed when it is in phase j is γ, for any j. A slight modification of the underlying phase process then allows us to compute the conditional age distribution of the individual at the first observation time, given that the individual is in phase $1 \leq j \leq m$ at that time. It suffices to add m absorbing phases $1', 2', \ldots, m'$ (one per transient phase), to the process so that phase j' is reached when the individual is observed in phase j. The initial distribution $\boldsymbol{\alpha}$ stays unchanged, but now the transition rate matrix becomes

$$\mathbf{T}(\gamma) = \mathbf{Q} - \gamma \mathbf{I},$$

and there are $m+1$ absorption rate vectors

$$\boldsymbol{q}_0, \boldsymbol{t}_{1'}(\gamma), \ldots, \boldsymbol{t}_{m'}(\gamma),$$

where $\boldsymbol{q}_0 = -\mathbf{Q}\mathbf{1}$ records the rates of absorption into phase 0 (corresponding to the death of the individual), and for $1' \leq j' \leq m'$, $\boldsymbol{t}_{j'}(\gamma) = \gamma \boldsymbol{e}_j$ records the rates of absorption into phase j' (corresponding to the observation of the individual in phase j). For the sake of clarity of the presentation, we shall drop the dependence on γ in $\mathbf{T}(\gamma)$ and $\boldsymbol{t}_{j'}(\gamma)$ and use the simpler notation $\mathbf{T}$ and $\boldsymbol{t}_{j'}$ in the sequel.

For any phase j (transient or absorbing), let

$$B(j) = \inf\{t \geq 0 : \varphi(t) = j\}$$

be first time the individual enters phase j, with $B(j) = \infty$ if the individual never enters phase j. Then, with probability one, precisely one of the random variables

$$\{B(0), B(1'), B(2'), ..., B(m')\}$$

is finite, and the age distribution conditional on the observed phase being j can be rewritten as

$$\mathrm{P}_{po}[A_o \leq x \,|\, \varphi_o = j] = \mathrm{P}[B(j') \leq x \,|\, B(j') < \infty]. \tag{4}$$

Based on this observation, the next proposition provides an expression for the conditional age distribution.

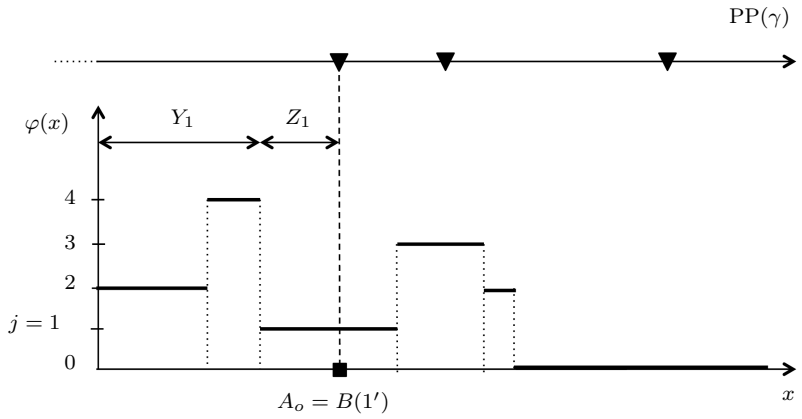


Fig. 1 A Possible trajectory of the phase process until absorption. The observation process is a Poisson process with rate γ (denoted as PP(γ)), where the ▼ symbols represent the events, and the phase is $j = 1$ at the first observation event.

Proposition 41 *The age distribution of the individual at the first observation (event) time of a Poisson(γ) process, conditional on the observed phase being j, is given by*

$$P_{po}[A_o \leq x \,|\, \varphi_o = j] = 1 - \frac{\boldsymbol{\alpha} e^{\mathbf{T}x} (-\mathbf{T})^{-1} \boldsymbol{e}_j}{\boldsymbol{\alpha}(-\mathbf{T})^{-1}\boldsymbol{e}_j}. \tag{5}$$

Proof. We have

$$\begin{aligned} \mathrm{P}[B(j') \leq x \,|\, B(j') < \infty] &= 1 - \mathrm{P}[B(j') > x \,|\, B(j') < \infty] \\ &= 1 - \frac{\mathrm{P}[x < B(j') < \infty]}{\mathrm{P}[B(j') < \infty]}. \end{aligned}$$

On the one hand,

$$\begin{aligned} \mathrm{P}[x < B(j') < \infty] &= \int_x^\infty \boldsymbol{\alpha} e^{\mathbf{T}u} \boldsymbol{t}_{j'} du \\ &= \boldsymbol{\alpha} e^{\mathbf{T}x} (-\mathbf{T})^{-1} \gamma \boldsymbol{e}_j, \end{aligned}$$

and on the other hand, since $\mathrm{P}[B(j') > 0] = 1$,

$$\mathrm{P}[B(j') < \infty] = \mathrm{P}[0 < B(j') < \infty] = \boldsymbol{\alpha} (-\mathbf{T})^{-1} \gamma \boldsymbol{e}_j, \tag{6}$$

which, with (4), completes the proof. □

The age at the observation time, conditional on the observed phase being j, can be written as the sum of two random variables,

$$A_o = Y_j + Z_j,$$

where Y_j denotes the last time that the Markov chain entered phase j before observation, and Z_j denotes the sojourn time in phase j between this time and observation, both random variables being conditionally independent given $\varphi_o = j$. The random variables A_o, Y_j, and Z_j are illustrated in Figure 1. Besides purely theoretical interest, the distribution of Y_j and Z_j may have practical interest when the phases have some physical interpretation (such as physiological ages), and an observer who sees an individual in phase j is interested in knowing the chronological age at which the individual entered that particular phase (Y_j), or for how long he/she has been in that phase (Z_j). The respective distributions of Y_j and Z_j are computed in the next two propositions.

Proposition 42 *The conditional distribution of Y_j, given $\varphi_o = j$, has a point mass at zero given by*

$$P_{po}[Y_j = 0 \,|\, \varphi_o = j] = \frac{\alpha_j}{(\lambda_j + \gamma)\,\boldsymbol{\alpha}(-\mathbf{T})^{-1}\boldsymbol{e}_j}, \tag{7}$$

and for $y > 0$,

$$P_{po}[Y_j \le y \,|\, \varphi_o = j] = 1 - \frac{\boldsymbol{\alpha} e^{\mathbf{T}y}(-\mathbf{T})^{-1}(\mathbf{Q} + \lambda_j\,\mathbf{I})\boldsymbol{e}_j}{(\lambda_j + \gamma)\,\boldsymbol{\alpha}(-\mathbf{T})^{-1}\boldsymbol{e}_j}, \tag{8}$$

where $\lambda_j = -Q_{jj}$.

Proof. First, recall from (6) that

$$\mathrm{P}[\varphi_o = j] = \mathrm{P}[B(j') < \infty] = \boldsymbol{\alpha}\,(-\mathbf{T})^{-1}\gamma\boldsymbol{e}_j.$$

Let T_o be the time at which the individual is observed, the clock being set at the individual's birth time. Thanks to the memoryless property of exponential random variables, T_o has the same distribution as the interarrival time in the Poisson observation process, that is, T_o is exponentially distributed with parameter γ. Then, for any $y > 0$, by conditioning on the value of T_o, we have

$$\begin{aligned}
&\mathrm{P}[Y_j \in [y, y+dy],\, \varphi_o = j] \\
&= \int_0^\infty \mathrm{P}[Y_j \in [y, y+dy],\, \varphi(u) = j \,|\, T_o \in [u, u+du]]\,\gamma e^{-\gamma u} du \\
&= \int_y^\infty \sum_{k \neq j} (\boldsymbol{\alpha} e^{\mathbf{Q}y})_k Q_{kj} e^{-\lambda_j (u-y)}\, dy\, \gamma e^{-\gamma u} du.
\end{aligned}$$

Next, observe that since $Q_{jj} = -\lambda_j$, we have

$$\sum_{k \neq j} (\boldsymbol{\alpha} e^{\mathbf{Q}y})_k Q_{kj} = \boldsymbol{\alpha} e^{\mathbf{Q}y}(\mathbf{Q} + \lambda_j\,\mathbf{I})\boldsymbol{e}_j.$$

As

$$\int_y^\infty e^{-\lambda_j (u-y)}\,\gamma e^{-\gamma u} du = \frac{\gamma e^{-\gamma y}}{\lambda_j + \gamma},$$

and $\mathbf{T} = \mathbf{Q} - \gamma\mathbf{I}$, we have

$$\mathrm{P}[Y_j \in [y, y+dy],\, \varphi_o = j] = \frac{\boldsymbol{\alpha} e^{\mathbf{T}y}(\mathbf{Q} + \lambda_j \mathbf{I})\boldsymbol{e}_j \gamma\, dy}{\lambda_j + \gamma}. \tag{9}$$

Similarly,

$$\mathrm{P}[Y_j = 0,\, \varphi_o = j] = \int_0^\infty \alpha_j e^{-\lambda_j u}\, \gamma e^{-\gamma u} du = \frac{\alpha_j \gamma}{\lambda_j + \gamma},$$

which leads to (7). Finally, (8) follows from (7) and (9) since

$$\mathrm{P}[Y_j \le y \,|\, \varphi_o = j] = \mathrm{P}[Y_j = 0 \,|\, \varphi_o = j] + \int_0^y \mathrm{P}[Y_j \in [u, u+du],\, \varphi_o = j],$$

for any $y > 0$. □

Proposition 43 *The conditional distribution of Z_j, given $\varphi_o = j$, is exponential with parameter $\gamma + \lambda_j$.*

Proof. Let S_j denote the sojourn time of the underlying Markov chain in phase j, and let $E_{j\to j'}$ denote the event that upon leaving phase j, the chain moves to phase j'. We have $\mathrm{P}[E_{j\to j'}] = \gamma/(\lambda_j + \gamma)$, and

$$\mathrm{P}[S_j > z, E_{j\to j'}] = e^{-(\lambda_j+\gamma)z} \frac{\gamma}{\lambda_j + \gamma},$$

therefore

$$\mathrm{P}[Z_j > z \,|\, \varphi_o = j] = \frac{\mathrm{P}[S_j > z, E_{j\to j'}]}{\mathrm{P}[E_{j\to j'}]} = e^{-(\lambda_j+\gamma)z},$$

which proves the statement of the proposition. □

4.2 Multiple observations

In practice, a multiple observation scheme would record a schedule of observation times $t_1, t_2, \ldots$, and observed phases $j_1, j_2, \ldots$. In that case, the only randomness in the individual's age would lie in the unknown first time t_1 of observation as the inter-observation times in the rest of the sample path are known. So the age distribution at the time of the last observation can be obtained directly from the age distribution at the time of the first observation. Furthermore the Markov property ensures that the sample path after time t_1 can tell us no more about the sample path before t_1 than the knowledge of phase j_1 at t_1.

In this section, we consider the possibly less realistic scenario where we know the observed phases but not the observation times, beyond the fact that they are points of a Poisson process. In this case, the distribution of the age at the last observation depends on all the observed phases. More precisely, we

assume that the observer makes $k \geq 2$ observations of an individual according to a Poisson process with rate γ. We further assume that the individual is still living at the time of the last (kth) observation. We shall compute the age distribution at the time of the kth observation, given the sequence of phases observed at the observation times.

For that purpose, we consider the same process as in the single observation case, with absorbing phases $0, 1', \ldots, m'$. Every observation event corresponds to a phase absorption in one of the phases $1', \ldots, m'$. After a phase absorption in j' (which corresponds to an observation of phase j), the process instantaneously starts again in phase j, that is, with the initial distribution vector $\boldsymbol{e}_j^\top$, until the next absorption event.

In order to properly define the quantities of interest, we need to redefine the random variables $B(j)$ as follows: for any initial phase distribution $\boldsymbol{\theta}$ and for any phase j,

$$B_{\boldsymbol{\theta}}(j) = \inf\{t \geq 0 : \varphi(t) = j, \varphi(0) \sim \boldsymbol{\theta}\} \tag{10}$$

is the first time the process reaches phase j, given that the initial phase follows the distribution $\boldsymbol{\theta}$. For the sake of clarity, we shall write $B_\ell(j)$ instead of $B_{\boldsymbol{e}_\ell^\top}(j)$ when the process starts in phase ℓ with probability one. Let $j_1, \ldots, j_k$ be the k successive observed phases. The age of the individual at the last observation time conditional on the observed phases, denoted by $A_o(j_1, \ldots, j_k)$ (or by A_o when there is no confusion), is then given by

$$A_o(j_1, \ldots, j_k) = B_{\boldsymbol{\alpha}}(j_1') + \sum_{i=2}^{k} B_{j_{(i-1)}}(j_i').$$

The age distribution at the time of the last observation conditional on the sequence of observed phases can then be written as

$$\mathrm{P}_{po}[A_o \leq x \,|\, B_{\boldsymbol{\alpha}}(j_1') < \infty, B_{j_1}(j_2') < \infty, \ldots, B_{j_{k-1}}(j_k') < \infty].$$

In order to compute this distribution, we need the following lemma, which is a particular case of Theorem 1 in [3]:

Lemma 44 *For $2 \leq i \leq k$, define the $m \times m$ matrix*

$$\mathcal{B}_{1i}(x) = \int_{u_1=0}^{x} \int_{u_2=0}^{x-u_1} \cdots \int_{u_{i-1}=0}^{x-u_1-\ldots-u_{i-2}} e^{\mathbf{A}_{11}u_1} \mathbf{A}_{12}\, e^{\mathbf{A}_{22}u_2} \mathbf{A}_{23} \ldots \mathbf{A}_{(i-1)i}\, e^{\mathbf{A}_{ii}(x-u_1-\ldots-u_{i-1})} d\boldsymbol{u},$$

where $x > 0$ and $\mathbf{A}_{ij}$ are constant $m \times m$ matrices. If the $km \times km$ block-structured matrix $\mathbf{A}^{(k)}$ is defined by

$$\mathbf{A}^{(k)} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \mathbf{A}_{23} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & & \ddots & & & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{A}_{(k-1)(k-1)} & \mathbf{A}_{(k-1)k} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{kk} \end{bmatrix},$$

then

$$\mathcal{B}_{1i}(x) = (\boldsymbol{f}_{k,1}^{\top} \otimes \mathbf{I}_m) e^{\mathbf{A}^{(k)}x} (\boldsymbol{f}_{k,i} \otimes \mathbf{I}_m), \tag{11}$$

where $\boldsymbol{f}_{k,i}$ *is a* $k \times 1$ *unit vector such that* $(\boldsymbol{f}_{k,i})_j = \delta_{ij}$. □

For the purpose of computing the conditional distribution of A_o, we define the $km \times km$ matrix $\mathbf{A}^{(k)}$ for any $k \geq 2$ as

$$\mathbf{A}^{(k)} = \begin{bmatrix} \mathbf{T} & \boldsymbol{e}_{j_1'}\boldsymbol{e}_{j_1}^{\top} & \mathbf{0} & \mathbf{0} \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{T} & \boldsymbol{e}_{j_2'}\boldsymbol{e}_{j_2}^{\top} & \mathbf{0} \ldots & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0}\ \mathbf{T} & \boldsymbol{e}_{j_{k-1}'}\boldsymbol{e}_{j_{k-1}}^{\top} \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0}\ \mathbf{0} & \mathbf{T} \end{bmatrix}. \tag{12}$$

Proposition 45 *For an arbitrary* $k \geq 2$*, the age distribution of the individual at the* k*th observation time, conditional on the successive observed phases being* $j_1, j_2, \ldots, j_k$*, is given by*

$$P_{po}[A_o \leq x \,|\, B_{\boldsymbol{\alpha}}(j_1') < \infty, B_{j_1}(j_2') < \infty, \ldots, B_{j_{k-1}}(j_k') < \infty] = \frac{N_k(x)}{D_k}, \tag{13}$$

where

$$\begin{aligned} N_k(x) &= \boldsymbol{\alpha}(I - e^{\mathbf{T}x})(-\mathbf{T})^{-1}\boldsymbol{e}_{j_1'} \prod_{i=1}^{k-1} \boldsymbol{e}_{j_i}^{\top}(-\mathbf{T})^{-1}\boldsymbol{e}_{j_{i+1}'} \\ &\quad - \boldsymbol{\alpha} \sum_{i=2}^{k} \mathcal{B}_{1i}(x)(-\mathbf{T})^{-1}\boldsymbol{e}_{j_i'} \prod_{\ell=i}^{k-1} \boldsymbol{e}_{j_\ell}^{\top}(-\mathbf{T})^{-1}\boldsymbol{e}_{j_{\ell+1}'}, \end{aligned} \tag{14}$$

$$D_k = \boldsymbol{\alpha}(-\mathbf{T})^{-1}\boldsymbol{e}_{j_1'} \prod_{i=1}^{k-1} \boldsymbol{e}_{j_i}^{\top}(-\mathbf{T})^{-1}\boldsymbol{e}_{j_{i+1}'}, \tag{15}$$

where $\mathcal{B}_{1i}(s)$ *is defined in* (11) *and* $\mathbf{A}^{(k)}$ *is given in* (12)*.*

The proof of Proposition 45 is given in the Appendix.

Using (11), the expressions for $N_k(x)$ and D_k can be rewritten as

$$N_k(x) = \boldsymbol{\alpha}\boldsymbol{u}_k - \boldsymbol{\alpha}e^{\mathbf{T}x}\boldsymbol{u}_k - \boldsymbol{v}_k e^{\mathbf{A}^{(k)}x}\boldsymbol{w}_k, \tag{16}$$

$$D_k = \boldsymbol{\alpha}\boldsymbol{u}_k \tag{17}$$

where

$$\boldsymbol{u}_k = (-\mathbf{T})^{-1}\boldsymbol{e}_{j_1'} \prod_{i=1}^{k-1} \boldsymbol{e}_{j_i}^{\top}(-\mathbf{T})^{-1}\boldsymbol{e}_{j_{i+1}'} \tag{18}$$

$$\boldsymbol{v}_k = (\boldsymbol{f}_{k,1}^{\top} \otimes \boldsymbol{\alpha}) \tag{19}$$

$$\boldsymbol{w}_k = \sum_{i=2}^{k} (\boldsymbol{f}_{k,i} \otimes I_m)(-\mathbf{T})^{-1}\boldsymbol{e}_{j_i'} \prod_{\ell=i}^{k-1} \boldsymbol{e}_{j_\ell}^{\top}(-\mathbf{T})^{-1}\boldsymbol{e}_{j_{\ell+1}'}.$$

In the expression for $\boldsymbol{w}_k$, an empty product (when $i = k$) is interpreted as the scalar 1. Note that it is also possible to express $N(k)$ and $D(k)$ recursively as follows: for $k \geq 3$,

$$N_k(x) = [N_{k-1}(x)\boldsymbol{e}_{j_{k-1}} - (\boldsymbol{f}_{k,1}^\top \otimes \boldsymbol{\alpha})e^{\mathbf{A}^{(k)}x}(\boldsymbol{f}_{k,k} \otimes I_m)](-\mathbf{T})^{-1}\boldsymbol{t}_{j'_k}$$
$$D_k = D_{k-1}\boldsymbol{e}_{j_{k-1}}(-\mathbf{T})^{-1}\boldsymbol{t}_{j'_k},$$

where $N_2(x)$ and D_2 are given in (38) and (37) respectively.

We now assume that the individual is still alive at the time of the kth observation, but is discovered dead (that is, in phase 0) at the time of the $(k+1)$st observation. We are then interested in the conditional lifetime distribution of the individual, given the sequence of observed phases. Indeed, the lifetime L it is then given by the age at the kth observation plus the time until absorption from the last observed phase j_k to phase 0, conditional on this time being less than the time between the kth and the $(k+1)$st observation. We shall need the following lemma:

Lemma 46 *Let $X \sim PH(\boldsymbol{\theta}, T)$ and $Y \sim Exp(\gamma)$. The conditional distribution of X, given that $X \leq Y$, is given by*

$$P[X \leq x \,|\, X \leq Y] = \frac{1 - \gamma\boldsymbol{\theta}(\gamma\mathbf{I} - \mathbf{T})^{-1}\mathbf{1} + \boldsymbol{\theta}e^{(\mathbf{T}-\gamma\mathbf{I})x}(\gamma\mathbf{I} - \mathbf{T})^{-1}\mathbf{T}\mathbf{1}}{1 - \gamma\,\boldsymbol{\theta}(\gamma\mathbf{I} - \mathbf{T})^{-1}\mathbf{1}},$$

and the density is given by

$$f_{X|X\leq Y}(x) = \frac{\boldsymbol{\theta}e^{(\mathbf{T}-\gamma\mathbf{I})x}(-\mathbf{T})\mathbf{1}}{1 - \gamma\boldsymbol{\theta}(\gamma\mathbf{I} - \mathbf{T})^{-1}\mathbf{1}}.$$

Proof. The distribution is obtained by conditioning on the value of Y. The expression for the conditional density then follows. □

The conditional lifetime distribution is computed in the next proposition.

Proposition 47 *The lifetime distribution of an individual, conditional on the sequence of observed phases being $j_1, j_2, \ldots, j_k, 0$, is given by*

$$\begin{aligned} &P_{po}[L \leq x \,|\, B_{\boldsymbol{\alpha}}(j'_1) < \infty, \ldots, B_{j_{k-1}}(j'_k) < \infty, B_{j_k}(0) < \infty] \\ &= C_k \left\{ \boldsymbol{\alpha}\boldsymbol{u}_k\boldsymbol{e}_{j_k}^\top[e^{(\mathbf{T}-\gamma\mathbf{I})x} - \mathbf{I}](\gamma\mathbf{I} - \mathbf{T})^{-1} + \boldsymbol{\alpha}\mathcal{I}_{12}(x) + \boldsymbol{v}_k\mathcal{J}_{1k}(x) \right\} \mathbf{T}\mathbf{1}, \end{aligned} \tag{20}$$

where $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$ are given in (18) *and* (19) *respectively, and*

$$C_k = [\boldsymbol{\alpha}\boldsymbol{u}_k(1 - \gamma\boldsymbol{e}_{j_k}^\top(\gamma\mathbf{I} - \mathbf{T})^{-1}\mathbf{1})]^{-1}, \tag{21}$$
$$\mathcal{I}_{12}(x) = (\boldsymbol{f}_{2,1}^\top \otimes \mathbf{I}_m)e^{\mathbf{B}^{(k)}x}(\boldsymbol{f}_{2,2} \otimes \mathbf{I}_m),$$
$$\mathcal{J}_{1k}(x) = [\mathbf{I}_{mk}, \mathbf{0}_{mk\times m}]e^{\mathbf{C}^{(k)}x}[\mathbf{0}_{m\times mk}, \mathbf{I}_m]^\top$$

with

$$\mathbf{B}^{(k)} = \begin{bmatrix} \mathbf{T} & \boldsymbol{u}_k\boldsymbol{e}_{j_k}^\top \\ \mathbf{0} & \mathbf{T} - \gamma\mathbf{I} \end{bmatrix}, \quad \mathbf{C}^{(k)} = \begin{bmatrix} \mathbf{A}^{(k)} & \boldsymbol{w}_k\boldsymbol{e}_{j_k}^\top \\ \mathbf{0} & \mathbf{T} - \gamma\mathbf{I} \end{bmatrix}.$$

Proof. We have

$$\begin{aligned}&\mathrm{P}[L \le x \,|\, B_{\boldsymbol{\alpha}}(j_1') < \infty, \ldots, B_{j_{k-1}}(j_k') < \infty, B_{j_k}(0) < \infty] \\ &= \mathrm{P}[(B_{\boldsymbol{\alpha}}(j_1') + \sum_{i=2}^{k} B_{j_{(i-1)}}(j_i')) + X \le x \,|\, B_{\boldsymbol{\alpha}}(j_1') < \infty, \ldots, B_{j_k}(0) < \infty] \quad (22)\end{aligned}$$

where $X \sim \mathrm{PH}(\boldsymbol{e}_{j_k}^\top, \mathbf{T})$ and X is taken conditionally on $X \le Y$, where Y is the interarrival time in the Poisson observation process, which is exponentially distributed with parameter γ. Since the age at the kth observation, given by $B_{\boldsymbol{\alpha}}(j_1') + \sum_{i=2}^{k} B_{j_{(i-1)}}(j_i')$, and the residual life time, X, are conditionally independent given the phase at the kth observation, j_k, we can use (13)–(15), together with Lemma 46 and the convolution formula, to compute the conditional lifetime distribution. In order to simplify the notation we define C_k as in (21). By conditioning on the value of X and using (16), we then obtain

$$\begin{aligned}&\mathrm{P}[L \le x \,|\, B_{\boldsymbol{\alpha}}(j_1') < \infty, \ldots, B_{j_{k-1}}(j_k') < \infty, B_{j_k}(0) < \infty] \\ &= \int_0^x \frac{N_k(x-u)}{D_k} \frac{\boldsymbol{e}_{j_k}^\top e^{(\mathbf{T}-\gamma\mathbf{I})u}\,(-\mathbf{T})\mathbf{1}}{(1-\gamma \boldsymbol{e}_{j_k}^\top(\gamma I - \mathbf{T})^{-1}\mathbf{1})} du \\ &= C_k \left\{ \boldsymbol{\alpha}\boldsymbol{u}_k \boldsymbol{e}_{j_k}^\top [\mathbf{I} - e^{(\mathbf{T}-\gamma\mathbf{I})x}](\gamma\mathbf{I}-\mathbf{T})^{-1} - \boldsymbol{\alpha}\int_0^x e^{\mathbf{T}(x-u)} \boldsymbol{u}_k \boldsymbol{e}_{j_k}^\top e^{(\mathbf{T}-\gamma\mathbf{I})u} du \right. \\ &\quad \left. -\boldsymbol{v}_k \int_0^x e^{\mathbf{A}^{(k)}(x-u)} \boldsymbol{w}_k \boldsymbol{e}_{j_k}^\top e^{(\mathbf{T}-\gamma\mathbf{I})u} du \right\} (-\mathbf{T})\mathbf{1} \\ &= C_k \left\{ \boldsymbol{\alpha}\boldsymbol{u}_k \boldsymbol{e}_{j_k}^\top [I - e^{(\mathbf{T}-\gamma\mathbf{I})x}](\gamma\mathbf{I}-\mathbf{T})^{-1} - \boldsymbol{\alpha}\mathcal{I}_{12}(x) - \boldsymbol{v}_k \mathcal{J}_{1k}(x) \right\} (-\mathbf{T})\mathbf{1},\end{aligned}$$

where, by Lemma 44,

$$\mathcal{I}_{12}(x) = (\boldsymbol{f}_{2,1}^\top \otimes I_m) e^{\mathbf{B}^{(k)}x} (\boldsymbol{f}_{2,2} \otimes I_m), \quad \mathcal{J}_{1k}(x) = [\mathbf{I}_{mk}, \mathbf{0}_{mk\times m}] e^{\mathbf{C}^{(k)}x} [\mathbf{0}_{m\times mk}, \mathbf{I}_m]^\top$$

with

$$\mathbf{B}^{(k)} = \begin{bmatrix} \mathbf{T} & \boldsymbol{u}_k \boldsymbol{e}_{j_k}^\top \\ \mathbf{0} & \mathbf{T}-\gamma\mathbf{I} \end{bmatrix}, \quad \mathbf{C}^{(k)} = \begin{bmatrix} \mathbf{A}^{(k)} & \boldsymbol{w}_k \boldsymbol{e}_{j_k}^\top \\ \mathbf{0} & \mathbf{T}-\gamma\mathbf{I} \end{bmatrix}.$$

□

5 The uniform observation scheme

In this section, we assume that an observer samples an individual in a population at a single random time T_o in accordance with a uniform distribution on $[0, t]$, for some time $t > 0$, where the clock is set at the birth of the individual. We then ask the same questions as in the Poisson observation scheme, but we expect different answers. The three random variables of interest A_o, Y_j, and Z_j are illustrated in Figure 2. Their respective conditional distribution, under the uniform observation scheme, is provided in the next three propositions.

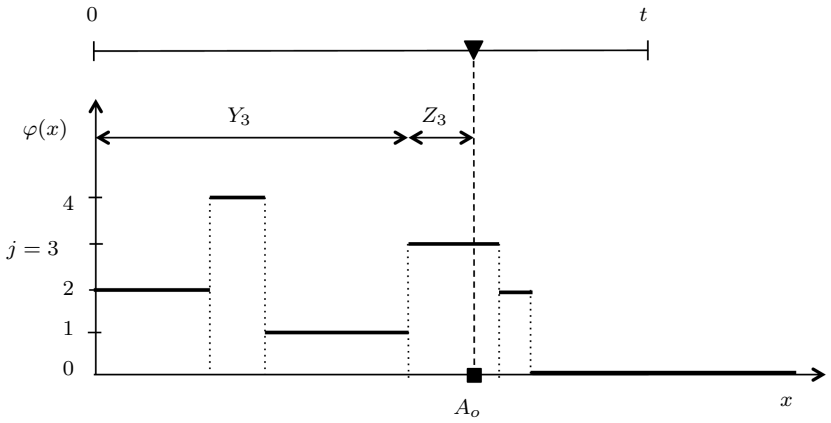


Fig. 2 A Possible trajectory of the phase process until absorption. The observation time is uniform on $[0,t]$ and is represented by a ▼ symbol. The observed phase is $j=3$.

Proposition 51 *The conditional age distribution of the individual at a random observation time uniformly distributed on* $[0,t]$*, given* $\varphi_o = j$*, is given by*

$$P_{uo}[A_o \le x \,|\, \varphi_o = j] = \begin{cases} 1 - \dfrac{\boldsymbol{\alpha}(e^{\mathbf{Q}x} - e^{\mathbf{Q}t})\,(-\mathbf{Q})^{-1}\boldsymbol{e}_j}{\boldsymbol{\alpha}(\mathbf{I} - e^{\mathbf{Q}t})(-\mathbf{Q})^{-1}\boldsymbol{e}_j} & \text{for } x \le t, \\ 1 & \text{for } x > t. \end{cases} \tag{23}$$

Proof. By conditioning on the value of the observation time T_o, we have

$$\mathrm{P}[\varphi_o = j] = (1/t)\int_0^t \boldsymbol{\alpha} e^{\mathbf{Q}u}\boldsymbol{e}_j\,du = (1/t)\,\boldsymbol{\alpha}(\mathbf{I} - e^{\mathbf{Q}t})(-\mathbf{Q})^{-1}\boldsymbol{e}_j, \tag{24}$$

and

$$\begin{aligned} \mathrm{P}[A_o \le x, \varphi_o = j] &= (1/t)\int_0^t \boldsymbol{\alpha} e^{\mathbf{Q}u}\boldsymbol{e}_j \mathbb{1}_{\{u\le x\}}du \\ &= (1/t)\int_0^{\min(x,t)} \boldsymbol{\alpha} e^{\mathbf{Q}u}\boldsymbol{e}_j du \\ &= (1/t)\,\boldsymbol{\alpha}(\mathbf{I} - e^{\mathbf{Q}\min(x,t)})\,(-\mathbf{Q})^{-1}\boldsymbol{e}_j. \end{aligned} \tag{25}$$

The conditional distribution given in (23) then follows by dividing (25) by (24), separating the case where $x \le t$ from that where $x > t$. □

Proposition 52 *The conditional distribution of* Y_j*, given* $\varphi_o = j$*, has a point mass at zero given by*

$$P_{uo}[Y_j = 0 \,|\, \varphi_o = j] = \frac{\alpha_j(1 - e^{-\lambda_j t})}{\lambda_j\,\boldsymbol{\alpha}(\mathbf{I} - e^{\mathbf{Q}t})(-\mathbf{Q})^{-1}\boldsymbol{e}_j}, \tag{26}$$

and for $0 < y \le t$,

$$P_{uo}[Y_j \le y \,|\, \varphi_o = j] = 1 - \frac{\lambda_j \boldsymbol{\alpha}(e^{\mathbf{Q}y} - e^{\mathbf{Q}t})(-\mathbf{Q})^{-1}\boldsymbol{e}_j + \boldsymbol{\alpha}e^{\mathbf{Q}y}\boldsymbol{e}_j(e^{-\lambda_j(t-y)} - 1)}{\lambda_j\, \boldsymbol{\alpha}(\mathbf{I} - e^{\mathbf{Q}t})(-\mathbf{Q})^{-1}\boldsymbol{e}_j}, \tag{27}$$

where $\lambda_j = -Q_{jj}$. *Finally,* $P_{uo}[Y_j \le y \,|\, \varphi_o = j] = 1$ *for* $y > t$.

Proof. The proof follows exactly the same lines as in the Poisson observation case. For any $0 < y \le t$,

$$\begin{aligned}
&\mathrm{P}[Y_j \in [y, y+dy],\, \varphi_o = j] \\
&= (1/t)\int_0^t \mathrm{P}[Y_j \in [y, y+dy],\, \varphi(u) = j \,|\, T_o \in [u, u+du]]\, du \\
&= (1/t)\int_y^t \sum_{k\neq j} (\boldsymbol{\alpha}e^{\mathbf{Q}y})_k Q_{kj}dy\, e^{-\lambda_j(u-y)}\, du \\
&= (1/t)\boldsymbol{\alpha}e^{\mathbf{Q}y}(\mathbf{Q} + \lambda_j\mathbf{I})\boldsymbol{e}_j\, dy \int_y^t e^{-\lambda_j(u-y)}\, du \\
&= (1/t)\boldsymbol{\alpha}e^{\mathbf{Q}y}(\mathbf{Q} + \lambda_j\mathbf{I})\boldsymbol{e}_j\, dy\, \frac{1 - e^{-\lambda_j(t-y)}}{\lambda_j}.
\end{aligned} \tag{28}$$

Similarly,

$$\mathrm{P}[Y_j = 0,\, \varphi_o = j] = (1/t)\int_y^t \alpha_j e^{-\lambda_j(u-y)}\, du = (1/t)\,\frac{\alpha_j(1 - e^{-\lambda_j(t-y)})}{\lambda_j},$$

which, together with (24), leads to (26). Finally, from (26), (28), and (24) we obtain (27). □

Proposition 53 *The conditional distribution of* Z_j, *given* $\varphi_o = j$, *is given by*

$$P_{uo}[Z_j \le z \,|\, \varphi_o = j] = 1 - \frac{\boldsymbol{\alpha}(\mathbf{I} - e^{\mathbf{Q}(t-z)})\,(-\mathbf{Q})^{-1}\boldsymbol{e}_j\, e^{-\lambda_j z}}{\boldsymbol{\alpha}(\mathbf{I} - e^{\mathbf{Q}t})(-\mathbf{Q})^{-1}\boldsymbol{e}_j} \quad \text{for } z \le t, \tag{29}$$

and $P_{uo}[Z_j \le z \,|\, \varphi_o = j] = 1$ *for* $z > t$.

Proof. By the usual arguments,

$$\begin{aligned}
&\mathrm{P}[Z_j \in [z, z+dz],\, \varphi_o = j] \\
&= (1/t)\left[\alpha_j e^{-\lambda_j z} + \int_z^t \sum_{k\neq j} (\boldsymbol{\alpha}e^{\mathbf{Q}(u-z)})_k Q_{kj} e^{-\lambda_j z}\, du\right] dz,
\end{aligned}$$

where the first term in the bracket accounts for the case where "$T_o = z$", that is, the individual is observed in her/his initial phase. We then have

$$\begin{aligned}
&\mathrm{P}[Z_j \in [z, z+dz],\, \varphi_o = j] \\
&= (1/t)\left[\alpha_j e^{-\lambda_j z} + \boldsymbol{\alpha}\int_z^t e^{\mathbf{Q}(u-z)}du\,(\mathbf{Q}+\lambda_j\mathbf{I})\boldsymbol{e}_j e^{-\lambda_j z}\right]dz \\
&= (1/t)\left[\alpha_j e^{-\lambda_j z} + \boldsymbol{\alpha}[\mathbf{I}-e^{\mathbf{Q}(t-z)}](-\mathbf{Q})^{-1}\,(\mathbf{Q}+\lambda_j\mathbf{I})\,\boldsymbol{e}_j e^{-\lambda_j z}\right]dz, \quad (30)
\end{aligned}$$

and the conditional density function of Z_j, given $\varphi_o = j$, is obtained by dividing (30) by (24) and rearranging the terms in the numerator:

$$f_{Z_j|\varphi_o=j}(z) = \frac{\boldsymbol{\alpha}\left[\lambda_j e^{-\lambda_j z} - e^{\mathbf{Q}t}e^{-(\mathbf{Q}+\lambda_j\mathbf{I})z}(\mathbf{Q}+\lambda_j\mathbf{I})\right](-\mathbf{Q})^{-1}\,\boldsymbol{e}_j}{\boldsymbol{\alpha}[\mathbf{I}-e^{\mathbf{Q}t}](-\mathbf{Q})^{-1}\boldsymbol{e}_j}.$$

Then, as $\mathrm{P}[Z_j \le z \,|\, \varphi_o = j] = \int_0^z f_{Z_j|\varphi_o=j}(u)\,du$, we obtain (29). □

6 Rare observation limit

In practice, individuals of an animal population are usually observed very seldom. This is particularly true for endangered wild populations such as the Chatham Island black robins *Petroica traversi*, which are observed once or twice per year (per individual) on average. We are therefore interested in the limit of the conditional age distribution as $\gamma \to 0$ in the Poisson observation scheme, or as $t \to \infty$ in the uniform observation scheme. First, observe that

$$\mathbf{T} \xrightarrow{\gamma\to 0} \mathbf{Q}, \quad \text{and} \quad e^{\mathbf{Q}t} \xrightarrow{t\to\infty} \mathbf{0}.$$

An interesting consequence of Propositions 41 and 51 is that the rare observation limit of the age distribution is identical for the Poisson and uniform observation schemes,

$$\lim_{\gamma\to 0} \mathrm{P}_{po}[A_o \le x \,|\, \varphi_o = j] = \lim_{t\to\infty} \mathrm{P}_{uo}[A_o \le x \,|\, \varphi_o = j].$$

In addition, the limiting age distribution corresponds to the age distribution $\mathrm{P}_{pb}[A_o \le x \,|\, \varphi_o = j]$ that we derived in Lemma 31 under the assumption that the birth process is Poisson.

Corollary 61 *The rare observation limit of the conditional age distribution, given $\varphi_o = j$, corresponds to $P_{pb}[A_o \le x \,|\, \varphi_o = j]$ whose expression is given by* (3). □

Actually, not only is the rare observation limit of the age distribution identical for the Poisson and the uniform observation schemes, but this holds for the conditional distribution of Y_j and Z_j too, as shown in the next two corollaries. These results are direct consequences of Propositions 42 and 52, and Propositions 43 and 53, respectively.

Corollary 62 *The rare observation limit of the conditional distribution of Y_j, given $\varphi_o = j$, has a point mass at zero and is given by*

$$P[Y_j = 0 \,|\, \varphi_o = j] = \frac{\alpha_j}{\lambda_j \, \boldsymbol{\alpha}(-\mathbf{Q})^{-1}\boldsymbol{e}_j} \tag{31}$$

and for $y > 0$,

$$P[Y_j \leq y \,|\, \varphi_o = j] = 1 - \frac{\boldsymbol{\alpha} e^{\mathbf{Q}y}(-\mathbf{Q})^{-1}(\mathbf{Q} + \lambda_j \, \mathbf{I})\boldsymbol{e}_j}{\lambda_j \, \boldsymbol{\alpha}(-\mathbf{Q})^{-1}\boldsymbol{e}_j}, \tag{32}$$

where $\lambda_j = -Q_{jj}$. □

Corollary 63 *The rare observation limit of the conditional distribution of Z_j, given $\varphi_o = j$, is exponential with parameter λ_j.* □

7 Numerical illustrations

We illustrate the results of the previous sections on a toy example first, and then on the real-world example of the Chatham Island black robin *Petroica traversi* population.

7.1 Toy example with five phases

We consider a $\mathrm{PH}(\boldsymbol{\alpha}, Q)$ lifetime distribution with $m = 5$ transient phases and transition rate matrix

$$Q = \begin{bmatrix} -3 & 2 & & & \\ & -5 & 3 & & \\ 1 & & -4 & 2 & \\ & 1 & & -6 & 3 \\ & & 1 & & -2 \end{bmatrix},$$

where a blank space represents a zero entry. We shall consider two initial distribution vectors:

$$\boldsymbol{\alpha}^{(1)} = [1, 0, 0, 0, 0], \quad \text{and} \quad \boldsymbol{\alpha}^{(2)} = [1/5, 1/5, 1/5, 1/5, 1/5].$$

In the first case, the process starts in phase 1 almost surely, while in the second case, the initial phase is chosen uniformly on the transient phase space. As we show below, the initial distribution vector can affect the shape of the various conditional distributions. We choose to represent densities (rather than distribution functions) as they better capture the features of the distributions.

We first assume a single observation, and we condition on the observed phase being $j = 4$. Figure 3 shows the conditional age densities obtained under the different observation schemes, for different values of the parameters γ and t, as well as the rare observation limit. Observe the discontinuity of the density at $x = t$ in the uniform case (while the distribution function given in (23) is

continuous at $x = t$). This comes from the fact that an individual observed in the time window $[0, t]$ cannot be older than t, and suggests that the Poisson observation scheme is more natural than the uniform observation scheme on a finite time interval. We see that for $\boldsymbol{\alpha}^{(1)}$, the mode of the distribution is positive and tends to increase as the observation becomes rare, while for $\boldsymbol{\alpha}^{(2)}$, the mode is clearly at age zero.

Figures 4 and 5 illustrate the conditional densities of Y_4 and Z_4 respectively. The initial distribution affects the shape of the density of Y_4, which has a point mass at zero for $\boldsymbol{\alpha}^{(2)}$, but has negligible effect on the shape of the density of Z_4. Also note that in the uniform observation case, the density of Y_j is continuous at $x = t$, but similar to the age density in Figure 3, the density of Z_j is discontinuous at $x = t$.

Finally, we consider the multiple Poisson observation scheme with $k = 5$ observations, for different values of the parameter γ, and different sequences of observed phases: $1, 2, 3, 3, 4$ (Sequence 1) and $2, 3, 1, 1, 4$ (Sequence 2, which is less likely than Sequence 1). In Figure 6 we compare the age distribution at the 5th observation, and the lifetime distribution given that the individual is dead at the 6th observation, for $\boldsymbol{\alpha}^{(1)}$ (there is not much difference for $\boldsymbol{\alpha}^{(2)}$). The graph illustrates how a change in the sequence of observed phases affects the related conditional distributions. We see that the tail of the distributions corresponding to Sequence 2 is fatter than for Sequence 1, that is, an individual is more likely to be older at the time of the last observation when Sequence 2 is observed.

In a last experiment, we compare the rare observation limit of the age distribution with the empirical distribution computed by simulating 5000 trajectories of the PH distribution and assuming that we observe the phases just before the next transition (that is, at the end of the exponential sojourn time interval). The result is illustrated in Figure 7 for $j = 4$ and for the two initial distributions $\boldsymbol{\alpha}$. We clearly see that the two distributions coincide, which supports the use of the rare observation limiting distribution in practical applications.

7.2 How old are the Chatham Island black robins *Petroica traversi*?

In this last section, we come back to our original objective, and illustrate the usefulness of our results to compute the age pyramid for the black robin population during the monitoring period between 2007 and 2014.

A first step of the analysis consists of modelling the female bird population using a branching process called the Markovian binary tree (MBT), which is described in detail in [7]. In an MBT, the lifetime of individuals is phase-type distributed, and individuals reproduce during their lifetime with a rate dependent on their current phase. Age-specific mortality and fertility rates of the black robins are estimated from the unique dataset collected between 2007 and 2014. They are shown for the first eleven age-classes in Table 1, where the total fertility rates have been divided by two to account for female

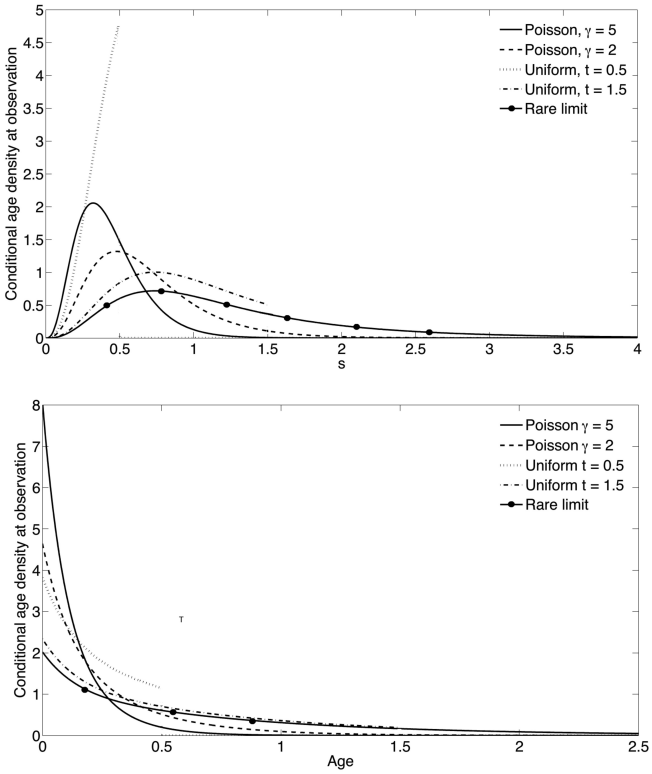


Fig. 3 Conditional age density at the observation time, given that phase $j = 4$ is observed, for the two observation schemes with different parameter values, and the rare observation limit, as $\boldsymbol{\alpha} = [1, 0, 0, 0, 0]$ (top) and $\boldsymbol{\alpha} = [1/5, 1/5, 1/5, 1/5, 1/5]$ (bottom).

offspring only, given a 50:50 sex-ratio on Rangatira Island where these data were collected [4]. As we describe below, these age-specific rates are used to estimate the parameters of an MBT that optimally fit the data, and this model can then be used to study demographic properties of the population.

In this particular case, the PH$(\boldsymbol{\alpha}, Q)$ lifetime distribution is assumed to have the following Coxian structure

$$\boldsymbol{\alpha} = \overset{1\ 2\ \ldots\ m}{\begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix}}, \quad Q = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ m \end{array} \overset{\begin{array}{ccccc} 1 & 2 & 3 & \ldots & m \end{array}}{\begin{bmatrix} -\lambda_1 & \lambda_1 s_1 & & & \\ & -\lambda_2 & \lambda_2 s_2 & & \\ & & & \ddots & \\ & & & & -\lambda_m \end{bmatrix}}; \tag{33}$$

that is, an individual starts its life in phase 1 and moves through successive phases until it dies; in this case, a transition from a transient phase j can only be to the next phase $j + 1$ (with probability s_j), or to the absorbing phase

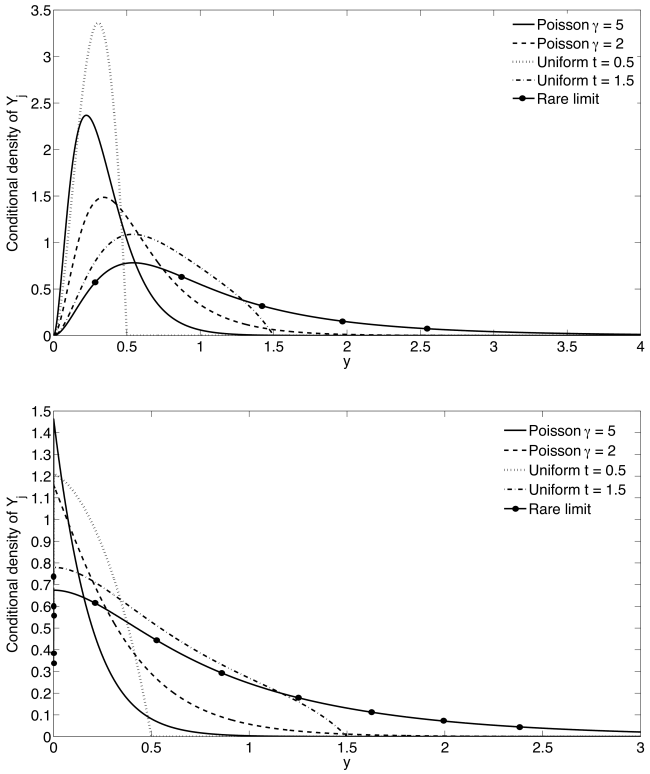


Fig. 4 Conditional density of Y_j, given that phase $j = 4$ is observed, for the two observation schemes with different parameter values, and the rare observation limit, as $\boldsymbol{\alpha} = [1, 0, 0, 0, 0]$ (top) and $\boldsymbol{\alpha} = [1/5, 1/5, 1/5, 1/5, 1/5]$ (bottom). The point mass at zero is clear when $\boldsymbol{\alpha} = [1/5, 1/5, 1/5, 1/5, 1/5]$.

0 (with probabilty $1 - s_j$). This particular PH distribution has a minimal number of parameters $(2m-1)$, which makes it more convenient for parameter estimation. In addition to the lifetime parameters, an additional $m \times 1$ vector $\boldsymbol{b} = [b_1, \ldots, b_m]^\top$ records the phase-specific birth rates of each individual in the MBT.

The model parameters are estimated following a weighted nonlinear regression approach based on the estimated age-specific mortality rates $\hat{d}_x$ and fertility rates $\hat{b}_x$ provided in Table 1. To summarize, the method consists of minimising the sum of weighted squared errors

$$F = \sum_{x=0}^{11} \left[(\hat{d}_x - \bar{d}(x))^2 + (\hat{b}_x - \bar{b}(x))^2 \right] \hat{S}_x, \tag{34}$$

where $\hat{S}_x = (1 - \hat{d}_0)(1 - \hat{d}_1) \cdots (1 - \hat{d}_{x-1})$ is the estimated probability of survival until age-class $[x, x + 1)$, and the functions $\bar{d}(x)$ and $\bar{b}(x)$ are the

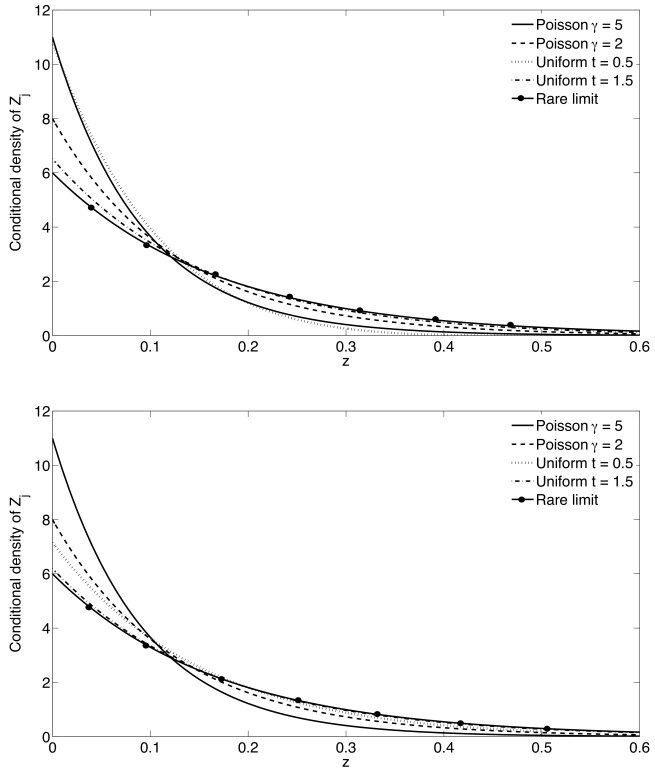


Fig. 5 Conditional density of Z_j, given that phase $j = 4$ is observed, for the two observation schemes with different parameter values, and the rare observation limit, as $\boldsymbol{\alpha} = [1, 0, 0, 0, 0]$ (top) and $\boldsymbol{\alpha} = [1/5, 1/5, 1/5, 1/5, 1/5]$ (bottom).

model equivalent to $\hat{d}_x$ and $\hat{b}_x$, respectively, and are given by

$$\bar{d}(x) = \frac{\boldsymbol{\alpha} e^{Qx}(I - e^{Q})\mathbf{1}}{\boldsymbol{\alpha} e^{Qx}\mathbf{1}}$$
$$\bar{b}(x) = \frac{\boldsymbol{\alpha} e^{Qx}(I - e^{Q})(-Q)^{-1}\boldsymbol{b}}{\boldsymbol{\alpha} e^{Qx}\mathbf{1}};$$

we refer to [7, Section 3] for the details.

Statistical validation tests conclude that taking $m = 8$ phases in the model provides a good balance between goodness of fit and complexity (see [7, Section 5]). The resulting optimal rates λ_i, b_i, and probabilities s_i, are provided in Table 2. Figure 8 shows the estimated age-specific mortality and fertility rates $\hat{d}_x$ (stars) together with the mortality and fertility functions $\bar{d}(x)$ and $\bar{b}(x)$ corresponding to the optimal MBT model (plain line). We see that the model curves smooth the point estimates in a satisfactory way.

Among other useful properties, the MBT model allows us to compute the asymptotic phase frequency in the population, that is, the proportion of birds

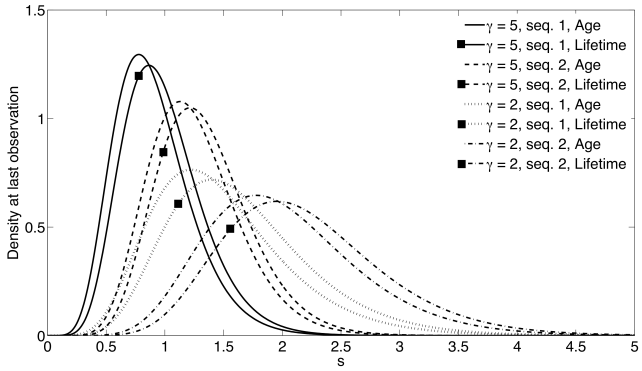


Fig. 6 Conditional age/lifetime density at the time of the last Poisson observation when the observed sequence is $1, 2, 3, 3, 4$ (Sequence 1) and $2, 3, 1, 1, 4$ (Sequence 2), with $\boldsymbol{\alpha} = [1, 0, 0, 0, 0]$.

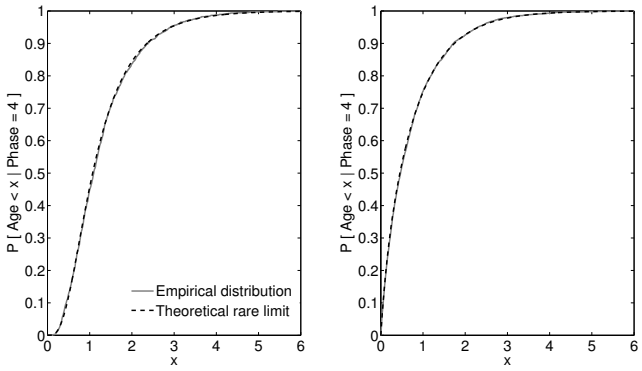


Fig. 7 . Comparison of the rare observation limit of the age distribution with the empirical distribution computed by simulations, for $\boldsymbol{\alpha} = [1, 0, 0, 0, 0]$ (right) and $\boldsymbol{\alpha} = [1/5, 1/5, 1/5, 1/5, 1/5]$ (left), when the observed phase is $j = 4$.

in each of the 8 phases if we let the population evolve for a long period of time with the same demographic rates. The asymptotic frequency of phase j, denoted by f_j^p, is given by

$$f_j^p = \frac{v_i}{\boldsymbol{v}\mathbf{1}},$$

where $\boldsymbol{v}$ is the left Perron-Frobenius eigenvector of the matrix $\Omega := Q - \boldsymbol{\alpha} \otimes \boldsymbol{b}$ [6]. The values f_j^p, $1 \leq j \leq 8$, are shown in Figure 9. Since the phases do not have any physical interpretation, the asymptotic phase frequency does not have much biological interest in its own right. However, it can be used in combination with the results developed in this paper to compute the asymptotic

Age-class $[x, x+1)$	Mortality rate $\hat{d}_x$	Fertility rate $\hat{b}_x$
$[0, 1)$	0.70	0.00
$[1, 2)$	0.16	1.15
$[2, 3)$	0.15	1.29
$[3, 4)$	0.17	1.46
$[4, 5)$	0.09	1.30
$[5, 6)$	0.29	1.00
$[6, 7)$	0.23	1.10
$[7, 8)$	0.25	1.38
$[8, 9)$	0.60	0.63
$[9, 10)$	0.33	0.50
$[10, 11)$	0.33	0.50

Table 1 Female age-specific mortality and fertility rates estimated from the raw data on the Chatham Island black robin population during the monitoring period 2007-2014. The rates are defined as follows: $\hat{d}_x$ is the probability that a female who reached age x dies within the year, and $\hat{b}_x$ is the expected number of female offspring per year from a female of age x.

Phase i	λ_i	s_i	b_i
1	4.79	0.30	0.01
2	0.94	0.92	0.01
3	1.96	0.91	0.08
4	1.53	0.88	4.53
5	1.03	0.89	0.47
6	0.95	0.90	1.40
7	0.98	0.91	2.83
8	0.69	-	0.15

Table 2 The parameters of the MBT modelling the female Chatham Island black robin population.

age-frequency (also called the age-pyramid), which cannot be obtained directly from the MBT model.

We consider the following eleven age-classes: $[0, 1)$, $[1, 2)$,...,$[9, 10)$, $[10, +)$, and we denote the asymptotic frequency of the age-class starting at age x as f_x^a, $0 \leq x \leq 10$. We can then compute f_x^a as

$$f_x^a = \sum_{1 \leq j \leq 8} f_j^p \, \mathrm{P}[\text{age} \in [x, x+1) \,|\, \text{phase} = j]. \tag{35}$$

We approximate the probability $\mathrm{P}[\text{age} \in [x, x+1) \,|\, \text{phase} = j]$ using the rare limit conditional age distribution provided in Corollary 61.

The resulting age pyramid for the black robins is depicted in Figure 10. From its shape, we see that a considerable number of offspring do not survive their first year in life, and although some individuals are known to live up to 16 years of age in this species, only a small fraction of individuals reach an age beyond 5 years. By comparison, Figure 11 shows the real age frequency in the female population in 2014. Even if the number of data per year is relatively small (there were only 93 female adults and 120 "female" eggs laid in 2014), the two pyramids exhibit similar shapes.

Given that the current black robin population is restricted to only two small islands and includes fewer than 300 individuals, the species remains

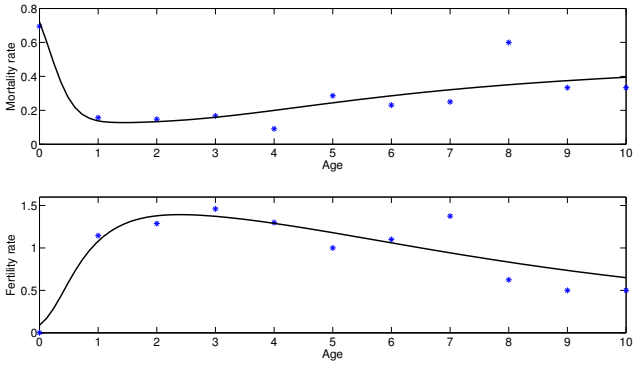


Fig. 8 Estimated age-specific mortality and fertility rates (stars), and the age-specific mortality and fertility curves computed with the optimal Markovian model fitting the data (plain lines).

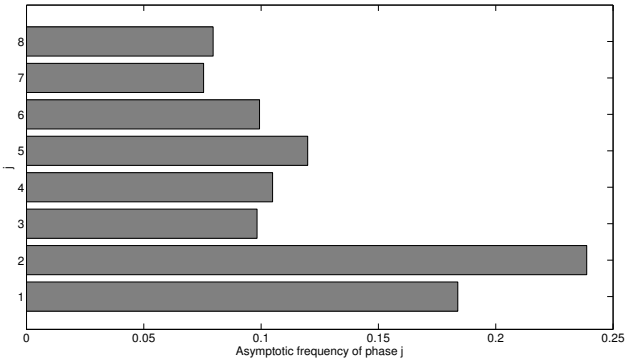


Fig. 9 The asymptotic phase frequency in the MBT modelling the Chatham Island black robin population, if the population evolves with the demographic rates of 2007-2014.

endangered [11]. Hence, knowing the age frequency of this population, and the associated fertility and mortality rates of different aged birds, is highly relevant to the future conservation management of this species.

Remark 71 *In the particular case of an ageing process with structure* (33)*, and conditionally on phase j being observed, any trajectory of the phase process before observation is restricted to the phases $1, 2, \ldots, j$. Therefore, for $j < m$, it is sufficient to consider the process restricted to the smaller phase space*

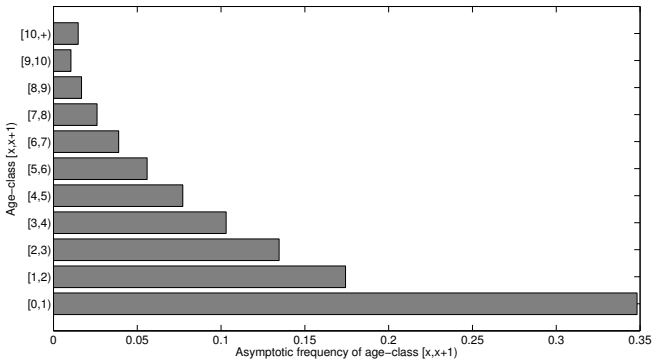


Fig. 10 The asymptotic age frequency in the Chatham Island black robin population if the population evolves with the demographic rates of 2007-2014.

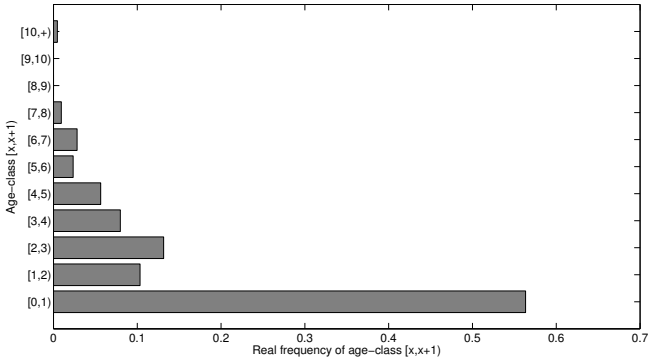


Fig. 11 The real age frequency in the female Chatham Island black robin population in 2014.

$\{1, 2, \ldots, j\}$, *with initial distribution vector and generator*

$$\boldsymbol{\alpha}^{(j)} = \overset{1\ 2\ \ldots\ j}{\begin{bmatrix} 1\ 0 \ldots 0 \end{bmatrix}},\ Q^{(j)} = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ j \end{array} \overset{\begin{array}{ccccc} 1 & 2 & 3 & \ldots & j \end{array}}{\begin{bmatrix} -\lambda_1 & \lambda_1 s_1 & & & \\ & -\lambda_2 & \lambda_2 s_2 & & \\ & & & \ddots & \\ & & & & -\lambda_j \end{bmatrix}}.$$

We further define the matrix $T^{(j)} = Q^{(j)} - \gamma I$ *where the identity matrix is* $j \times j$, *and* $\boldsymbol{e}_i^{(j)}$ *as the truncation of* $\boldsymbol{e}_i$ *after its* j*th entry, for* $i \leq j$. *The matrix ingredients* $\boldsymbol{\alpha}, Q, T$, *and* $\boldsymbol{e}_j$ *used in the lemmas, propositions, and corollaries in the previous sections can then be replaced by their smaller counterpart* $\boldsymbol{\alpha}^{(j)}$, $Q^{(j)}$, $T^{(j)}$, *and* $\boldsymbol{e}_j^{(j)}$.

Finally, observe that with the particular ageing structure (33)*, in the Poisson observation scheme, the random variable Y_j has the same distribution as the time until the process with generator $T^{(j)}$ reaches phase j for the first time, $B(j)$, conditionally on $B(j) < \infty$.*

Appendix: Proof of Proposition 45

We have

$$\begin{aligned}
&\mathrm{P}[A_o \le x \,|\, B_{\boldsymbol{\alpha}}(j_1') < \infty, B_{j_1}(j_2') < \infty, \ldots, B_{j_{k-1}}(j_k') < \infty] \\
&= \frac{\mathrm{P}[A_o \le x,\, B_{\boldsymbol{\alpha}}(j_1') < \infty, \ldots, B_{j_{k-1}}(j_k') < \infty]}{\mathrm{P}[B_{\boldsymbol{\alpha}}(j_1') < \infty, B_{j_1}(j_2') < \infty, \ldots, B_{j_{k-1}}(j_k') < \infty]} \\
&=: \frac{\bar{N}_k(x)}{\bar{D}_k}.
\end{aligned}$$

We shall prove using induction on k that

$$\frac{\bar{N}_k(x)}{\bar{D}_k} = \frac{N_k(x)}{D_k}, \tag{36}$$

where $N_k(x)$ and D_k satisfy (14) and (15), respectively. Recall that $\boldsymbol{t}_{j'} = \gamma \boldsymbol{e}_j$ for any absorbing phase $1' \le j' \le m'$. When $k = 2$,

$$\begin{aligned}
\bar{D}_2 &= \mathrm{P}[B_{\boldsymbol{\alpha}}(j_1') < \infty, B_{j_1}(j_2') < \infty] \\
&= \boldsymbol{\alpha}(-\mathbf{T})^{-1}\boldsymbol{t}_{j_1'}\, \boldsymbol{e}_{j_1}^{\top}(-\mathbf{T})^{-1}\boldsymbol{t}_{j_2'} \\
&= \gamma^2 D_2.
\end{aligned} \tag{37}$$

Further, by (10) and by conditioning on the value of the absorption times $B_{\boldsymbol{\alpha}}(j_1')$ and $B_{j_1}(j_2')$, we have

$$\begin{aligned}
\bar{N}_2(x) &= \mathrm{P}[(B_{\boldsymbol{\alpha}}(j_1') + B_{j_1}(j_2')) \le x,\, B_{\boldsymbol{\alpha}}(j_1') < \infty, B_{j_1}(j_2') < \infty] \\
&= \int_{u=0}^{x} \boldsymbol{\alpha} e^{\mathbf{T}u}\boldsymbol{t}_{j_1'} \int_{v=0}^{x-u} \boldsymbol{e}_{j_1}^{\top} e^{\mathbf{T}v}\boldsymbol{t}_{j_2'}\, dv\, du \\
&= \gamma^2 \int_{u=0}^{x} \boldsymbol{\alpha} e^{\mathbf{T}u}\boldsymbol{e}_{j_1'}\boldsymbol{e}_{j_1}^{\top}(\mathbf{I} - e^{\mathbf{T}(x-u)})(-\mathbf{T})^{-1}\boldsymbol{e}_{j_2'}\, du \\
&= \gamma^2\{\boldsymbol{\alpha}(\mathbf{I} - e^{\mathbf{T}x})(-\mathbf{T})^{-1}\boldsymbol{e}_{j_1'}\boldsymbol{e}_{j_1}^{\top}(-\mathbf{T})^{-1}\boldsymbol{e}_{j_2'} - \boldsymbol{\alpha}\mathcal{B}_{12}(x)(-\mathbf{T})^{-1}\boldsymbol{e}_{j_2'}\} \\
&= \gamma^2\, N_2(x),
\end{aligned} \tag{38}$$

where

$$\mathcal{B}_{12}(x) = \int_{u=0}^{x} e^{\mathbf{T}u}\boldsymbol{e}_{j_1'}\boldsymbol{e}_{j_1}^{\top} e^{\mathbf{T}(x-u)} du.$$

Using Lemma 44, this matrix integral can be evaluated explicitly by defining the $2m \times 2m$ block-structured matrix

$$\mathbf{A}^{(2)} = \begin{bmatrix} \mathbf{T} & \boldsymbol{e}_{j_1'}\boldsymbol{e}_{j_1}^{\top} \\ \mathbf{0} & \mathbf{T} \end{bmatrix},$$

so that

$$\mathcal{B}_{12}(x) = (\boldsymbol{f}_{2,1}^\top \otimes \mathbf{I}_m) e^{\mathbf{A}^{(2)}x} (\boldsymbol{f}_{2,2} \otimes \mathbf{I}_m).$$

Therefore (36) holds for $k = 2$.

We now assume that (36) holds for k, and we need to prove that is still holds for $k+1$. We can decompose the conditional age at the $(k+1)$st observation, $A_o(j_1, \ldots, j_{k+1})$, into the sum of the random variables $B_{\boldsymbol{\alpha}}(j_1')$ and $A_o(j_2, \ldots, j_{k+1})$, which are conditionally independent given j_1. Note that $A_o(j_2, \ldots, j_{k+1})$ is now conditional on the phase process starting with initial distribution vector $\boldsymbol{e}_{j_1}^\top$ rather than $\boldsymbol{\alpha}$, and the first observed phase is j_2 rather than j_1, etc. To avoid confusion, we shall use the notation $\hat{A}_o(j_2, \ldots, j_{k+1})$ (or $\hat{A}_o$ for short), $\hat{N}_k(x), \hat{D}_k, \hat{\boldsymbol{\mathcal{B}}}_{1i}(x)$ whenever we will be in that situation.

We use the convolution formula for the sum of the two conditionally independent variables $B_{\boldsymbol{\alpha}}(j_1')$ and $\hat{A}_o$, together with the conditional distribution of $B_{\boldsymbol{\alpha}}(j_1')$ given in (5) and the induction assumption, to obtain

$$\begin{aligned}
&\mathrm{P}[(B_{\boldsymbol{\alpha}}(j_{1'}) + \hat{A}_o) \le x \,|\, B_{\boldsymbol{\alpha}}(j_1') < \infty, B_{j_1}(j_2') < \infty, \ldots, B_{j_k}(j_{k+1}') < \infty] \\
&= \int_0^x \frac{\boldsymbol{\alpha} e^{\mathbf{T}u} \boldsymbol{t}_{j_1'}}{\boldsymbol{\alpha}(-\mathbf{T})^{-1} \boldsymbol{t}_{j_1'}} \mathrm{P}[\hat{A}_o \le x - u | B_{j_1}(j_2') < \infty, \ldots, B_{j_k}(j_{k+1}') < \infty] \, du \\
&= \int_0^x \frac{\boldsymbol{\alpha} e^{\mathbf{T}u} \boldsymbol{e}_{j_1'}}{\boldsymbol{\alpha}(-\mathbf{T})^{-1} \boldsymbol{e}_{j_1'}} \frac{\hat{N}_k(x-u)}{\hat{D}_k} du.
\end{aligned}$$

We immediately see that the denominator of the above expression, $D_{k+1} := \boldsymbol{\alpha}(-\mathbf{T})^{-1}\boldsymbol{e}_{j_1'}\hat{D}_k$, corresponds to (15) for $k+1$. It remains to show that the numerator, $N_{k+1}(x) := \int_0^x \boldsymbol{\alpha} e^{\mathbf{T}u} \boldsymbol{e}_{j_1'} \hat{N}_k(x-u) du$, corresponds to (14) for $k+1$. Using (14) and letting $\boldsymbol{r}_{i,k} = (-\mathbf{T})^{-1}\boldsymbol{t}_{j_{i+1}'} \prod_{\ell=i}^{k-1} \boldsymbol{e}_{j_{\ell+1}}^\top (-\mathbf{T})^{-1} \boldsymbol{t}_{j_{\ell+2}'}$, we have

$$\begin{aligned}
&\int_0^x \boldsymbol{\alpha} e^{\mathbf{T}u} \boldsymbol{t}_{j_1'} \hat{N}_k(x-u) \, du \\
&= \int_0^x \boldsymbol{\alpha} e^{\mathbf{T}u} \boldsymbol{t}_{j_1'} \left\{ \boldsymbol{e}_{j_1}^\top (\mathbf{I} - e^{\mathbf{T}(x-u)}) \boldsymbol{r}_{1,k} - \boldsymbol{e}_{j_1}^\top \sum_{i=2}^k \hat{\boldsymbol{\mathcal{B}}}_{1i}(x-u) \boldsymbol{r}_{i,k} \right\} du \\
&= \boldsymbol{\alpha}(\mathbf{I} - e^{\mathbf{T}x})(-\mathbf{T})^{-1} \boldsymbol{t}_{j_1'} \boldsymbol{e}_{j_1}^\top \boldsymbol{r}_{1,k} - \boldsymbol{\alpha} \boldsymbol{\mathcal{B}}_{12}(x) \boldsymbol{r}_{1,k} \\
&\quad - \boldsymbol{\alpha} \sum_{i=2}^k \int_0^x e^{\mathbf{T}u} \boldsymbol{t}_{j_1'} \boldsymbol{e}_{j_1}^\top \hat{\boldsymbol{\mathcal{B}}}_{1,i}(x-u) \, du \, \boldsymbol{r}_{i,k}.
\end{aligned}$$

Using Lemma 44 and (11), we can show that

$$\int_0^x e^{\mathbf{T}u} \boldsymbol{t}_{j_1'} \boldsymbol{e}_{j_1}^\top \hat{\boldsymbol{\mathcal{B}}}_{1,i}(x-u) \, du = \boldsymbol{\mathcal{B}}_{1,i+1}(x),$$

so that by properly redefining the indices we finally obtain what we need. □

Acknowledgements The authors are supported by the Australian Research Council Laureate Fellowship FL130100039. The first author has also conducted part of the work under

the Discovery Early Career Researcher Award DE150101044. Field-based research on black robins from 2007-14 was funded by a New Zealand Foundation for Research, Science and Technology fellowship to MM (UOCX0601), and the University of Canterbury, the Brian Mason Scientific and Technical Trust and the Mohamed bin Zayed Species Conservation Fund.

References

1. O. O. Aalen. Phase-type distributions in survival analysis. *Encyclopedia of Biostatistics* (1995).
2. D. Butler and D. Merton. *The Black Robin: Saving the World's Most Endangered Bird*. Oxford University Press, Auckland (1992).
3. F. Carbonell, J. C. Jimenez, and L. M. Pedroso. Computing multiple integrals involving matrix exponentials. *Journal of Computational and Applied Mathematics*, 213(1):300–305 (2008).
4. I. Cubrinovska, M. Massaro, and M. L. Hale. Assessment of hybridisation between the endangered Chatham Island black robin (*Petroica traversi*) and the Chatham Island tomtit (*Petroica macrocephala chathamensis*). *Conservation Genetics*, 17:259-265 (2016).
5. L. A. Gavrilov and N. S. Gavrilova. The biology of life span: a quantitative approach. (1991).
6. S. Hautphenne and G. Latouche. The Markovian binary tree applied to demography. *Journal of Mathematical Biology*, 64(7):1109–1135 (2012).
7. S. Hautphenne, M. Massaro, and K. Turner. Fitting Markovian binary trees using global and individual demographic data. Submitted for publication. Preprint available on *https://arxiv.org/abs/1702.04281*, (2017).
8. 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.
9. X. S. Lin and X. Liu. Markov aging process and phase-type law of mortality. *North American Actuarial Journal*, 11(4):92–109 (2007).
10. M. Massaro, R. Sainudiin, D. Merton, J. V. Briskie, A. M. Poole, and M. L. Hale. Human-assisted spread of a maladaptive behavior in a critically endangered bird. *PloS one*, 8(12):e79066, 2013.
11. 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.
12. M. F. Neuts. *Matrix-geometric solutions in stochastic models: an algorithmic approach*. Courier Dover Publications (1981).