

GALTON-WATSON PROCESS AND BAYESIAN INFERENCE: A TURNKEY METHOD FOR THE VIABILITY STUDY OF SMALL POPULATIONS

Bertrand Cloez, T Daufresne, Marion Kerioui, Bénédicte Fontez

▶ **To cite this version:**

Bertrand Cloez, T Daufresne, Marion Kerioui, Bénédicte Fontez. GALTON-WATSON PROCESS AND BAYESIAN INFERENCE: A TURNKEY METHOD FOR THE VIABILITY STUDY OF SMALL POPULATIONS. 2019. hal-01994949

HAL Id: hal-01994949
https://hal.science/hal-01994949

Preprint submitted on 25 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

GALTON-WATSON PROCESS AND BAYESIAN INFERENCE: A TURNKEY METHOD FOR THE VIABILITY STUDY OF SMALL POPULATIONS

B. CLOEZ, T. DAUFRESNE, M. KERIOUI, B. FONTEZ

ABSTRACT.
1 Sharp prediction of extinction times is needed in biodiversity monitoring and conservation management.
2 The Galton-Watson process is a classical stochastic model for describing population dynamics. Its evolution is like the matrix population model where offspring numbers are random. Extinction probability, extinction time, abundance are well known and given by explicit formulas. In contrast with the deterministic model, it can be applied to small populations.
3 Parameters of this model can be estimated through the Bayesian inference framework. This enables to consider non-arbitrary scenarios.
4 We show how coupling Bayesian inference with the Galton-Watson model provides several features: i) a flexible modelling approach with easily understandable parameters ii) compatibility with the classical matrix population model (Leslie type model) iii) A non-computational approach which then leads to more information with less computing iv) a non-arbitrary choice for scenarios, parameters... It can be seen to go one step further than the classical matrix population model for the viability problem.
5 To illustrate these features, we provide analysis details for two examples whose one of which is a real life example.

1. **Introduction** Population viability analysis (PVA) [7, 45, 8, 40, 38] aims to manage or predict population development from some ecological, genetic or demographic assumptions. It generally relies on modelling purposes. The simulations are frequently used to work on these questions even if some works highlight the use of Mathematical results for these problems as [13, 25, 30, 39, 19] among many others.

Habitat degradation is pointed out as the major cause of extinction for many threatened species around the world. Yet, for species that are exploited for food or other purposes (e.g., exploited fish populations, bushmeat etc...), or purposely destroyed (e. g., large predators), extinction often occurs in intact habitats. For these cases, the assumption that population dynamics takes place in a fixed habitat seems reasonable. Then, assuming a given environment and a short period of time without brutal variation of the environment, are we able to propose a model to support decision making for management?

In this paper, we will illustrate how to answer some questions emanating from conservation biology such as:

(I) What are the odds on saving a population? How probable is extinction? For how long can the population survive?
(II) How sharply are numbers expected to fall in the short term?
(III) How can we proceed to avoid extinction [14, 15], to maintain a population size [43] or wipe out some population [6, 33, 23]. For instance, if we are able to add some individuals in a population, how many re-introductions are needed to reach a threshold of variability for the population with high probability?

These simple questions, when the environment is assumed unchanged/stable are not currently being resolved successfully and we propose a reformulation of classical models that leads to understand the population viability with or without possible actions.

To illustrate this, Cairns, Ross and Taimre in [9] underlined that the usual class of diffusion models are simple but lead to inaccurate predictions of critical values such as the expected time to extinction. They advocated that a more appropriate model may be a discrete-state Markov process describing the actual number of individuals in the population like the birth–death processes. These models are not used in practice because they are difficult to work with, from both analytical and computational points of view.

Starting from the famous Matrix population [4, 32, 31], we will introduce the simplest birth-death process of Galton Watson which leads asymptotically to exponential growth and does not take into account some density-dependence dynamics. This assumption is relevant if we address population dynamics in populations at low density, whose dynamics is generally not subject to density dependence factors. On the other hand, these populations may exhibit small numbers, which may preclude the use

Date: Compiled January 25, 2019.

of deterministic models. We will show step by step how to use it for extinction prediction.

Although extensive literature exists on the frequentist estimation for the Galton-Watson processes (see for instance [1, 12, 18, 34] and references therein), we choose a Bayesian reasoning [41]. In contrast with a frequentist approach, it has the advantage that a prior with expert knowledge or previous analyses can be introduced in the models. This may be especially useful for small data sets. Indeed, data sets for rare and elusive species, which are often the target of PVA, are generally scarce and incomplete. Moreover, a posterior allows a precise quantification of the uncertainty due to the parameter estimations. This gives posterior probability estimates or credible intervals for the various parameters allowing to build natural scenarios without arbitrary choice. This is of particular interest for practitioners.

Besides, our approach is based on a remarkable relationship between the Galton-Watson processes and Bayesian inference. As we will see, the Galton-Watson processes have multinomial transitions and verify the Markov property, the law of the parameters conditioned according to the historical paths (namely the posterior) can have an explicit form. Indeed, choosing a classical conjugate prior of the multinomial distribution enables recovery of this distribution as a posterior (with different parameters) after one generation and then the Markov property enables to generalize this property for all generations.

This allows to use classical formulas on Galton-Watson processes and integrate it under the posterior to predict quantitatively the behaviour of the population; that is the growth, extinction, relative abundance...

This simple and powerful expression of the posterior is one of the main strengths of our approach. Indeed, this remarkable property seems to be observed only in [37] in the mono-type case (i.e. $K = 1$), and, to the best of our knowledge, has never been applied in practice with a mono-type nor with many types. We will show how this leads to a simple and powerful way to analyse the viability of the population.

Outline: In the two next Sections, Section 2 and Section 3, we present the Galton-Watson processes and our stastical model. Section 3 is more technical and contains several formulas and properties of such processes. Section 4 states the posterior on which we will apply the properties developed in Section 3. Section 5 is devoted to two pedagogical examples. the first is based on synthetic data and is concerned with extinction. The second concerns the French bear population and survival, which apart from illustrating our method, is of great interest in is own right. Finally Section 6 gives some useful simple extensions.

Box 0. Notation

- K is the number of types.
- $N(t) = (N_1(t), \dots, N_K(t))^\intercal$ designes the state vectors (sample size of each type) at time t.
- $p_{i,j}(k)$ is the probability that an individual of type i gives k individuals of type j after one unit of time. It is supposed to be a random number.
- $\kappa_{i,j}$ is the maximal number of offspring of type j from an individual of type i.
- $(\alpha_{i,j}(k))$ design the prior knowledge in terms of sample size of individuals in type i which gives k individuals of type j.
- $\xi_{i,j,l,t}$ is the (random) number of offspring of type j of the l-th individual of type j.
- $M = (M_{i,j}) = \left(\sum_{k=0}^{\kappa_{i,j}} k p_{i,j}(k)\right)$ is the random *mean* matrix.
- λ is the principal eigenvalue of M.
- u, v are respectively the left and right eigenvector of M.
- T_{Ext} is the (random) extinction time. Namely the duration after the present time T to have extinction.
- s_i is the probability of extinction starting from one individual with type i.
- $\mathbf{n} = (n_{i,j}(k,t))_{1\le i,j\le K, k\ge 0, T\ge t\ge 0}$ are the data. they correspond to the number of individuals of type i with k offspring of type j at time t for all types and times over $\{0, ..., T\}$.
- α is a statistical threshold. Classicaly $\alpha = 0.05$ or 0.1.

Box 1. Pathways to PVA through Bayesian-Galton-Watson approach

Here, we summarize some steps for carrying out a viability study using data. Several details are given in Appendix 8.2.

(1) **Previous knowledge**: Initialize your hyper parameters $(\alpha_{i,j}(k))$.

(2) Compute the following quantity of interest:

- **Short time evolution**: number of individuals:
$$\mathbb{E}[N(t) \mid \mathbf{n}] = \mathbb{E}[M^t \cdot N(0) \mid \mathbf{n}].$$
- **Probability of viability**:
$$\mathbb{P}(\lambda > 1 \mid \mathbf{n}).$$
If $\lambda < 1$ then reintroduction is useless, one has to act on the environment.
- **Time to extinction**: if $\lambda < 1$, extinction is certain and one can draw
$$U(t) : t \mapsto \mathbb{E}[\lambda^t \mid \mathbf{n}] \times \frac{\sum_{j=1}^K N_j(0) v_j}{\min_{1\le j\le K} v_j}.$$
Indeed when $U(t)$ is below a threshold α (as 1 or 5%) we know that after this time there is more than $1-\alpha$ that all individuals are extinct; that is $T_{\text{Ext}} \le t$. An upper bound for T_{Ext} is given in Section 8.2.
- **Probability of extinction**: if $\lambda > 1$ then the population has the capacity to survive but extinction can occur if the population is too small:
$$\mathbb{P}(\text{Ext}) = \mathbb{E}\left[s_1^{N_1(0)} \times \cdots \times s_K^{N_K(0)} \mid \mathbf{n}\right]$$
- **How to plan reintroduction**: draw the laws of $s_1, \ldots, s_K$ to see the survival potential of each type and therefore choose to reintroduce the more efficient one.

2. **Galton Watson process: a Stochastic Matrix Population Model** One of the most famous and powerful mathematical settings for PVA is the Matrix population model. They were introduced by [4, 32, 31] and they have become a classical tool to predict the evolving size of a population as illustrated in the books [13, 46]. In this model, a population is classified into discrete types and its evolution is in discrete time. The state of the population, at a certain time $t \in \mathbb{N} = \{0, 1, ...\}$, is given by a (column) vector, $N(t) = (N_1(t), \ldots, N_K(t))^\intercal$, where K is the number of types (and the exponent $\intercal$ designs the usual transposed vector) . This vector evolves from time t to $t+1$ with the following matrix product:

$$N(t+1) = A \cdot N(t),$$

that is, for every type j,

$$N_j(t+1) = \sum_{i=1}^K A_{i,j} N_i(t). \tag{1}$$

Matrix A is called the population projection matrix and the entry $A_{i,j}$ tells how many individuals of type i appear per individual of type j. Classically the types correspond to the age (or the location), and A then codes the development (or the migration), the birth and the death.

This classical model makes it possible to understand population variation and resulting extinction or viability. Indeed, let us consider for instance the one-type case $K = 1$. We then have

$$N_1(t+1) = A_{1,1} N_1(t) \Rightarrow N(t) = A_{1,1}^t N_1(0), \tag{2}$$

for all $t \ge 0$, which entails

$$\lim_{t\to\infty} N_1(t) = \begin{cases} 0 & \text{if } A_{1,1} < 1 \\ N_1(0) & \text{if } A_{1,1} = 1 \\ +\infty & \text{if } A_{1,1} > 1. \end{cases}$$

In particular when $A_{1,1} > 1$ the population is viable and when $A_{1,1} < 1$ the population goes into extinction. This result can easily be generalized with K types using the principal eigenvalue of A. Namely if λ is the largest eigenvalue of A (in modulus) and N the associated eigenvector then $N(t) \underset{t\to\infty}{\sim} \lambda^t N$. The vector N then relates the asymptotic relative abundance although λ represents the growth rate. See for instance [13] for details.

In case of extinction, a central question in PVA is to have an estimate of the extinction time. Nevertheless, any prediction of the extinction time is impossible since population size is a deterministic real number. Namely, we have $N_1(t) > 0$ for all $t \ge 0$ as soon as $N_1(0) > 0$. When $K = 1$ and $A_{1,1} < 1$, or more generally $\lambda < 1$, extinction means $\lim_{t\to\infty} N(t) = 0$ but at each time there are some individuals even if this number can be very small. Hence, mathematically the extinction time cannot be defined as the time T_{Ext} that $N_1(T_{\text{Ext}}) = 0$ for the matrix population model. The usual setting to quantify time to extinction in this setting is therefore the so-called quasi-extinction time [24, 28, 35, 17]. That is, we fix some arbitrary threshold γ under which the number of individuals will be insufficient to ensure persistence of the population, and the associated quasi-extinction time T_γ is defined as the moment the population size reaches the level γ. This threshold has to be chosen in order to take into account demographic stochasticity, Allee effect, *etc*. When $K = 1$, Equation (2) gives

$$T_\gamma = \frac{\log(\gamma) - \log(N_1(0))}{\log(A_{1,1})},$$

and we see that the problem is that this time crucially depends on the arbitrary choice of γ.

Another drawback to considering real numbers to model the population size is the unit choice. Indeed does $N_1(t) = 1$ mean that there is $1, 10, 10^2, 10^3, ...$ individuals? Then how can we use this type of model for small population sizes as in [10, 21] or [19, Section 5] for instance? Note that, as deterministic models are proved to be realistic only in large populations, this unit choice has to be large. But how many individuals is enough to be large depends on the model and its parameters. Even if deterministic models can give suitable results for only a hundred individuals, a unit choice of 10^3 (or even 10^6) may not be enough to avoid atto-fox type problem [11] (that is having 10^{-18} individuals able to avoid extinction).

As pointed out in [25, 30], to overcome these problems, it is more natural to consider the population size as a discrete number (namely an integer) as this is the case in Galton-Watson processes.

Multi-type Galton-Watson processes, were introduced in [5, 47], and can be seen as a matrix population model with offspring randomness (but with non-random environment). In this approach, at each time t, each individual l of type i gives a discrete random number $\xi_{i,j,l,t} \in \{0, 1, 2, \dots\}$ of child type j; namely formula (1) became

$$N_j(t+1) = \sum_{i=1}^{K} \sum_{l=1}^{N_i(t)} \xi_{i,j,l,t}. \tag{3}$$

Of course setting

$$A_{i,j}(t) = \frac{1}{N_i(t)} \sum_{l=1}^{N_i(t)} \xi_{i,j,l,t},$$

we recover the form of equation (1) that is

$$N(t+1) = A(t) \cdot N(t).$$

With this formalism, $A(\cdot)$ is a random matrix changing at each time. However it is not the same as in [46, Chapter 3] because this sequence of matrices is not a sequence of independent and identically distributed matrices. To see how demography or environment play a role in this matrix notation and to see the importance of considering demographic stochasticity (even in larger populations), we can see [19] and references therein.

Consistently, as in matrix populations, one can recover the asymptotic type distributions and the increasing rate of the population through decomposition of a matrix, but one can go further and find sharp results on extinction. The credit that can be given to these predictions can also be estimated. That is, we can give a precise confidence interval for the extinction risk, the extinction time, *etc.* This again is one of the main differences with matrix population models. The proof of these results can be found in [2, 29, 26, 27, 36]. Some of them have already been applied for real-life examples; see for instance [14] or [25, page 208]. However, besides using the mathematical properties of Galton-Watson processes, we also propose in this work an efficient and consistent statistical approach to using it. Indeed, as noticed in [39] "Conducting a population viability analysis involves the steps of choosing an appropriate model, fitting the model to data, and using the fitted model to predict the extinction risk ", the approach that we describe aims to cover all these PVA steps from some life tables or more generally from demographic data.

To conclude, one should keep in mind, that as with matrix population models, the Galton-Watson process crucially depends on the two following strong assumptions:

(1) interaction between individuals has no effect on the population dynamics;
(2) parameters of one individual depend on this individual only through its type. In particular, all individuals of the same type have the same distribution of offspring and this does not vary in time.

3. Our model: a Galton-Watson process embedded in a Bayesian framework

To define a multi-type Galton-Watson process, we need to introduce some probabilities $p_{i,j}$ on $\{0, 1, ...\}$. They will represent the progeny of type j from an individual of type i. More precisely, an individual oftype i will give, after one unit of time, k individuals of type j, with probability $p_{i,j}(k)$. We assume the existence of a maximal number of offspring $\kappa_{i,j}$; that is $p_{i,j}(k) = 0$, for all $k > \kappa_{i,j}$. We have then K^2 probabilities $p_{1,1}, p_{1,2}, ..., p_{K,K}$ which are represented in a vector form $p_{i,j} = (p_{i,j}(0), ..., p_{i,j}(\kappa_{i,j}))$.

For real life problems, these (fundamental) parameters $p_{i,j}$ are unknown. In a frequentist framework, these parameters of interest would have been assumed to be unknown, but fixed. Namely, it would be assumed that in the population, for all i, j there is only one true probability distribution $p_{i,j}$. In the Bayesian view of subjective probability, all of these unknowns are treated as uncertain and therefore should be described by a probability distribution over probability distribution.

We will then suppose that $(p_{i,j})_{1 \le i,j \le K}$ is distributed according to some probability distribution, usually called the prior. All through this paper, this prior will be the convolution of Dirichlet distribution [41, Section A.8 p. 521]; namely $p_{i,j}$ are independent from each other and probability-valued random variables whose law is given for every i, j by

$$p_{i,j} \sim \mathbf{Dir}\left(\alpha_{i,j}(0), ..., \alpha_{i,j}(\kappa_{i,j})\right),$$

for some sequence $(\alpha_{i,j}(k))_{k=0,\dots,\kappa_{i,j}}$. This is a classical law on discrete probability whose expression is given

in Appendix 8.1. Parameters $(\alpha_{i,j})$ are called hyper parameters and correspond to the information that you want to incorporate in your estimation. They are fixed by the users and can take into account previous studies or expert knowledge.

Among all the choices, two main strategies can be chosen.

The first one corresponds to the non-informative choice. If we have no information on the ecological system, we wish to decrease the effects of prior outcomes. We then take $p_{i,j}$ as drawn uniformly at random over all probability vectors. This corresponds to the choice $\alpha_{i,j}(0) = \cdots = \alpha_{i,j}(\kappa_{i,j}) = 1$. Namely, we consider as prior, the uniform law on the simplex of probability measure on a discrete space. This choice is motivated by the minimization of the entropy of these distributions (See [41, Section 3.2.3] and [42, Section 6]). Roughly, it is the more random choice or the less informative one.

Another choice can consist of putting information in the hyperparameters $(\alpha_{i,j})$. Indeed, as Dirichlet distribution is uni-modal (at least for large parameters) then one can easily choose $(\alpha_{i,j})$ so that $\mathbb{E}[p_{i,j}(k)]$ is equal to a presupposed valued m_k plus or minus a fixed error estimate σ (uniform in k). Values of m_k and σ can for instance be taken from a previous study where data are not available but just the estimator with a confidence interval as in [15, 14]. Then the expert choice is driven by external data and the weight of the expertize is automatic. Finally $(\alpha_{i,j})$ can incorporate belief which does not come from any data. These parameters also represent the balance between data and expertize.

Details for initializing $(\alpha_{i,j})$ in practice are given in Appendix 8.2.

Forthwith, a Galton-Watson process $N = (N(t))_{t=1,2,\dots}$ is a random vector-valued sequence satisfying equation (3). In this equation, conditionally on $(p_{i,j})$, random variables $\xi_{i,j,l,t}$ are supposed to be independent and $\xi_{i,j,l,t}$ are distributed according to $p_{i,j}$ on $\{0, 1, ..., \kappa_{i,j}\}$. Namely,

$$\mathbb{P}\big(\forall i,j,l,t,\ \xi_{i,j,l,t} \in A_{i,j,l,t} \mid (p_{i,j})\big) = \prod_{l,t\in\mathbb{N}} \prod_{i,j=1}^{K} \sum_{k\in A_{i,j,l,t}} p_{i,j}(k),$$

for all sets $A_{i,j,l,t} \subset \mathbb{N}$. This is a discrete-time model. Although there is a continuous-time version of the Galton-Watson processes and our result should hold for continuous times, we restrict ourself to a discrete-time setting because of the general nature of the data. Indeed data is often collected at punctual moments. Number $N(t)$ then represents the number of individuals at t years, t days or more relevant choices of times.

An important quantity associated with this model the random *mean* matrix M whose entries are

$$M_{i,j} = \sum_{k=0}^{\kappa_{i,j}} k p_{i,j}(k). \tag{4}$$

It is a random matrix since $p_{i,j}$ are supposed random (Bayesian framework). This matrix is the counterpart to the population projection matrix in matrix population model [13, 46]. When $p_{i,j}(0) = 1$ for $j \notin \{0, i, i+1\}$ that is one can just get old or reproduce then $M_{i,j} = 0$ for $j \notin \{0, i, i+1\}$ and we recover the classical Leslie matrix. Under general assumptions [26, Section 2.3.1], the evolution of $N(t)$ when t becomes large only depends on the largest eigenvalue λ of this matrix and the two associated eigenvectors u, v (with positive coordinates). That are those verifying

$$uM = \lambda M,\ Mv = \lambda v \tag{5}$$

and

$$\sum_{i=1}^{K} u_i v_i = 1,\ \sum_{i=1}^{K} u_i = 1$$

On the event $\lambda < 1$ then the process goes to extinction; that is there exists a finite random time T_{Ext} such that $N(t) = 0$ for all $t \geq T_{\text{Ext}}$ (and $N(T_{\text{Ext}} - 1) > 0$). Even if one can estimate T_{Ext} by simulation, sharp analytic bounds have been proved. For instance,

$$\begin{aligned} &\mathbb{P}(T_{\text{Ext}} > t \mid N_1(0), \dots, N_K(0), (p_{i,j})) \\ &= \mathbb{P}\left(\sum_{i=1}^{K} N_i(t) > 0 \mid N_1(0), \dots, N_K(0), (p_{i,j})\right) \\ &\leq \lambda^t \frac{\sum_{i=1}^{K} v_i N_i(0)}{\min_{1\leq i\leq K} v_i}. \end{aligned} \tag{6}$$

See for instance [26, Equation (5.54)]. This bound allows to find an upper bound of the extinction time with high probability (see Box 1 and Section 8.2). To our knowledge, there was no equivalent lower bound in great generality. However, [26, Box 5.2 p. 119] gives such a bound when $K = 1$. This allows to have a lower bound on the extinction time. In Appendix 8.2, we generalize this result for $K > 1$ and show how use it and (6) to bound the extinction time.

On the event $\lambda > 1$, extinction occurs with positive probability (but not equal to 1). In fact, if Ext denotes the extinction event then we have

$$\begin{aligned} &\mathbb{P}(\text{Ext} \mid N_1(0), \dots, N_K(0), (p_{i,j})) \\ &\quad = s_1^{N_1(0)} \times \cdots \times s_K^{N_K(0)}, \end{aligned} \tag{7}$$

where $s = (s_1, ..., s_K)$ is the unique solution of $s = \varphi(s)$ in the simplex $\triangle = \{s \in [0,1]^K \mid \sum_{i=1}^{K} s_i = 1\}$.

Function $\varphi = (\varphi_i)_{1\leq i\leq K}$ is the generating function associated with p; for $i \in \{1, ..., K\}$, it is defined by

$$\varphi_i(s_1, ..., s_K) = \sum_{i_1,...,i_k} s_1^{i_1} p_{i,1}^{i_1} \dots s_K^{i_K} p_{i,K}^{i_K}. \tag{8}$$

In particular, when we start with only one individual with type i (that is $N_i(0) = 1$ and $N_j(0) = 0$ for $j \neq i$) then

$$\mathbb{P}(\text{Ext} \mid N_1(0), \dots, N_K(0), (p_{i,j})) = s_i.$$

Additionally, Process $(N(t))$ is a Markov process. That means that for a current time T, the future evolution of the population only depends on the distribution of the population at time T and not on periods of time in the past (conditionally on $(p_{i,j})$ of course). As a consequence, Formula (6) Formula (7) hold when replacing $N(0)$ by $N(T)$ and conditioning on $N(T) > 0$. Know that some individuals remain at time T, we have to take into account this information.

Finally, in addition to the spectral analysis, all results of a matrix population model (sensitivity analysis,...) can be applied using the matrix M for the study of the Galton-Watson process. Also, for short time, the Galton-Watson process evolves in the same way as this associated deterministic model up to error around $\sqrt{N(t)}$.

All of these mathematical results can be found in [2, 29, 26, 27, 36].

4. Main result: explicit expression of the posterior

We assume that our demographic data has the form of life tables $(n_{i,j}(k,t))_{1\leq i,j\leq K,k,t\geq 0}$; where $n_{i,j}(k,t)$ corresponds to the number of individuals at time t of type i which gives k individuals of type j at time $t+1$. As for example in Table 1. We assume we know the population over the time interval $\{0, \dots, T\}$; namely our data is $\mathbf{n} = (n_{i,j}(k,t))_{1\leq i,j\leq K,k\geq 0,T\geq t\geq 0}$.

Due to the well known conjugation property between multinomial and Dirichlet distribution (see for instance [41, 42]), and the form of the Galton-Watson transitions, the form of the posterior is simple. More precisely, we can see this for every i, j, the posterior of the parameter $p_{i,j}$ conditioned on the data $\mathbf{n}$ are

$$\begin{aligned}\mathcal{L}(p_{i,j} \mid \mathbf{n}) \\ = \mathbf{Dir}\left(\alpha_{i,j}(0) + \sum_{t=0}^{T} n_{i,j}(0,t), \right. \\ \left. \dots, \alpha_{i,j}(\kappa_{i,j}) + \sum_{t=0}^{T} n_{i,j}(\kappa_{i,j}, t)\right).\end{aligned} \tag{9}$$

The proof of this formula is given in Appendix 8.1.

This posterior turns into an extensive summary of the information available on the parameter, integrating simultaneously modelling choices and available observations. To answer our main questions, we must summarize this object into the principal objects of interest (risk of extinction, extinction time, abundance, ...). Box 1 and Appendix 8.2 gives a typical "to-do" list in order to answer some main questions in PVA. We directly use the posterior to calculate precisely the probability of some events. For instance, one can calculate

$$\mathbb{P}(Ext \mid (n)),$$

directly integrating Equation (7) under the Dirichlet distribution given in Equation (9). This has the advantage of not being conditioned on one $(p_{i,j})$ because we take a mean value of all $(p_{i,j})$ combinations, using our posterior. This perfectly describes the prediction randomness coming from the demographic stochasticity and the statistical estimation.

One possible drawback is that all quantities of interest (abundance...) are random objects and may be difficult to understand or describe. Using the full posterior can then be difficult for practitioners and it can be helpful to sum up the information. We describe here two possible ways of aggregating the posterior according to the objectives of the outcomes: an estimator approach and a scenario approach.

Firstly, as explained in detail in [41, Section 4], the posterior mean, the posterior mode, the posterior median, $etc.$ can be used to build an estimator of the input of interest. If a choice must be made among the types of estimators above, there is no way of selecting a best estimator without using an error-type criterion; see [41, Section 4] for details. Note that some of these types of estimator can be easily computed; for instance

$$\mathbb{E}[M_{i,j} \mid \mathbf{n}] = \frac{\sum_{k=1}^{\kappa_{i,j}} k \left(\alpha_{i,j}(0) + \sum_{t=0}^{T} n_{i,j}(k,t)\right)}{\sum_{k=0}^{\kappa_{i,j}} \left(\alpha_{i,j}(0) + \sum_{t=0}^{T} n_{i,j}(k,t)\right)}. \tag{10}$$

The main drawback of this approach is that it loses the error estimation which is naturally embedded in the posterior. However, as soon as these values are fixed one can directly calculate λ (estimation of viability), probability of extinction s_1, s_2 ...

Finally, an intermediate approach between the full posterior approach and the punctual estimation consists of building scenarios from the posterior. In PVA, in order to make some decisions, it is (more or less) usual to build some scenarios for the future; see for instance [14, 15]. Nevertheless, in these papers, the variation of the main estimation to build the difference between the scenarios is not so clear. A natural choice for making scenarios that take into account the data and the error (due to the variance) that we make in the estimation is to consider

different quantiles of the posterior. For instance, from Table 2, we can choose for $p_{0,1}(1)$, the values 0.67, 0.82, 0.93 for scenarios that underestimate, estimate or overestimate the mortality rate (note that there is no symmetry).

5. **Examples** In this section, we give two examples. The first one based on simulated data aims to show in a pedagogical way our method and compare it with previous study when population goes into extinction. The second example is a real life example. It is concerned with the French Pyrenean brown bear whose viability has a major interest from an ecological and political point of view.

Synthetic data: comparison of methods Let us consider a simple population model going to extinction. We assume there is only one type $K = 1$ and the maximum number of offspring is $\kappa_{1,1} = 3$. We assume that survival and reproduction events are independent. All individuals survive with probability $p_S = 0.4$ and have k offspring with probability $p_R(k)$ with p_R given by the vector

$$p_R = (0.8, 0.1, 0.05, 0.05).$$

Namely each individual gives birth to 1 child with probability 0.1, 2 with probability 0.05 and 3 with probability 0.05. This is a Galton-Watson process with non-random parameters. With our previous notation, we have

$$\begin{aligned}
p^\star_{1,1}(0) &= (1-p_S)p_R(0) = 0.48\\
p^\star_{1,1}(1) &= p_S p_R(0) + (1-p_S)p_R(1) = 0.38\\
p^\star_{1,1}(2) &= p_S p_R(1) + (1-p_S)p_R(2) = 0.07\\
p^\star_{1,1}(3) &= p_S p_R(2) + (1-p_S)p_R(3) = 0.05\\
p^\star_{1,1}(4) &= p_S p_R(3) = 0.02.
\end{aligned}$$

The expected *mean* matrix $M^\star = \sum_{k=0}^{4} k p^\star_{1,1}(k) = 0.75$ is just a number and $\lambda^\star = M^\star = 0.75 < 1$ and then extinction is certain. We note $M^\star, \lambda^\star$, $p^\star_{1,1}$, $etc.$ for the true parameters instead of $M, \lambda, p_{1,1}$, $etc.$ because we keep this notation for the random variables distributed according to our prior which will serve to estimate $M^\star, \lambda^\star$, $p^\star_{1,1}$, $etc.$ Let us begin by simulating a population for 5 years to create a learning data set, that is our knowledge on the population. This is resumed in Table 1. Focus on this table, we recover that

$$N(t) = \sum_{k=0}^{4} n_{1,1}(k,t),\ N(t+1) = \sum_{k=0}^{4} k n_{1,1}(k,t).$$

For instance, at time 0 there are 100 individuals of whom 47 die without offspring, 39 survive without offspring or die with one offspring $etc.$ There are then 75 individuals at time 1 and so on.

Continuing the simulation we see that the population evolves as follows

$$N(T+1) = 19,\ N(T+2) = N(T+3) = 13,$$
$$N(T+4) = 8,\ N(T+5) = 5,$$
$$N(T+6) = 1,\ N(T+7) = 0.$$

Therefore $T^\star_{\text{Ext}} = 7$ in this sample. The aim of what follows is to give bounds on this quantity using only Table 1.

We choose a non-informative prior $\alpha_{1,1} = 1$. We then find that, conditionally on data of Table 1, $p_{1,1}$ is distributed according to a Dirichlet law with parameter $(145, 128, 20, 14, 8)$. Indeed it is the sum of each line of Table 1 plus $\alpha_{1,1} = 1$. We then find

$$\mathbb{E}[M \mid \mathbf{n}] = 0.7689, \quad \mathbb{P}(\lambda > 1 \mid \mathbf{n}) = 10^{-04}.$$

That is we find that the posterior mean estimator is relatively close to the true value $M^\star = 0.75$ and that the population is not viable with a probability very close to 1.

Now using the method developed in Appendix 8.2 with thresholds $\alpha = 0.05$ we find that the extinction time should be between 3 and 31; which is what happens because $T_{\text{Ext}} = 7$.

We compare our results with two other methods.

The first one is based on a diffusion model [22]. Using the method and notation of Foley [22], we find that the mean and variance of the growth rate are respectively $r_d = -0.3028$ and $v_r = 0.0041$ in log scale. We have $e^{r_d} = 0.7387$ which is relatively close to $M^\star = 0.75$. We then obtain a good parameter estimation. However to estimate the time to extinction, we have to choose a carrying capacity charge. At best, we obtain 36548 which is far from the real extinction time. This could be expected because we are not in the setting of [22] which is concerned with a more stable population (not decreasing one) whose extinction arises with rare events (designed as environmental stochasticity by Foley).

We also use a naive approach based on linear regression. Starting from Equation (2), we can fit the trend line:

$$\log(N(t+1)) = r_d + \log(N_t) = r_d t + c_d,$$

and then look at the time spent given by the upper and lower confidence lines (at 90% threshold) to reach 0. The lower bound is 7 and the upper bound is 9. These bounds are correct because they include the true value $T_{\text{Ext}} = 7$.

Taking only one simulation has only a qualitative or pedagogical interest. Let us repeat the previous steps 1000 times to see how many times the extinction is well predicted. With our method the extinction time is in our interval in around 93% of simulations instead of 49% from the naive regression. To end this example, we plot the density of the extinction through these 1000 simulations in Figure 1 and Figure 2. In Figure 1, we add the normal law with the same first two moments. We see that the distribution of the extinction time is not Gaussian at all: it is not symmetric and has a heavier tail. It is therefore not suitable to search for some values and build a

confidence interval through Gaussian quantiles. In Figure 2, we can illustrate the law of our lower and upper bound.

Real data: French Pyrenean brown bears The Pyrenean brown bear (Ursus arctos) population is considered as one of the most seriously threatened with extinction in Western Europe. In the 90's and early 2000s, the reinforcement of the population by the introduction of a few individuals from Slovenia allowed to avoid population extinction. Yet, the population remains fragile, and the introduction of new individuals gives rise to tumultuous debates in France. Let us give here a short study on the powerful properties of our approach (and the difference with the preceding example).

The data we used comes from [10]. It corresponds to the exhaustive supervision of all the population between 2005 and 2016 of the French Pyrenean brown bears. This population is split into two different isolated subpopulations. In the western part of the Pyrenees, before 2018, there were only two males (of which only one is indigenous), therefore we will focus on the subpopulation living in the central part of the Pyrenees.

We consider the same structure model as [15]. Namely, we only consider the evolution of the number of females and we consider a population structured by age, with $K = 5$ classes. For classes $i \in \{1, \dots, 4\}$, this corresponds to bears whose age is $i - 1$ years. For $i = 5$ this corresponds to the bears whose age is greater than or equal to 4 years. The life of a bear is modelled as follows: during its first 4 age stages, it can either die or pass to the next age. When it is in the last fully developed stage, it can die, survive and reproduce. See [15] for biological motivations.

We assume that surviving and reproducing are two independent events even if generally it is assumed that reproduction is possible only when the female parent survive. Both choices are modelization preferences that marginally add a bias to the prediction. It can have an importance mostly for the interpretation of the estimated parameters but we are not doing this here. Indeed, as we estimate our paramaters with this assumption, it decreases the values of reproducing parameters (because we will not see offspring when parents die) but it will increase the offspring predictions (because it is possible to have offspring when parents die). These two facts are balanced because they are assumed for the estimation and for the prediction. In the same way, we assume that females can reproduce each year although it is an exceptional event. This changes nothing because it will naturally divide by 2 the reproduction rate.

Let us now follow the steps of Box 1.

We start by choosing the information that we want to take into account. Begin by considering the non-informative setting, that is, for all $i \in \{1, ..., 4\}$ and $k \in \{0, 1\}$

$$\alpha_{i,i+1}(k) = 1,$$

and for $k \in \{0, ..., 3\}$

$$\alpha_{5,1}(k) = 1.$$

Using Formula (9), the posterior law is easily calculable from the data of [10] and is then given in Table 2. The Beta distribution can be seen as a particular case of Dirichlet distribution.

However, we only have 11 years of data and one would think that this is not sufficient to estimate all the parameters we need. Let us show how we can integrate the statistical estimation of [15]. The authors used data coming from different articles with not necessarily the full description of the estimation procedure. If we look at the infantile mortality rate $p_{1,2}(0)$ then it was based on [48]. Their estimation is 0.4 for a population of 150 bear cubs. The resulting estimated variance is then $1.96\sqrt{0.4 \times 0.6/150} = 0.0784$. To take into account this study, we can then choose

$$\alpha_{1,2}(0) = 1.2417, \qquad \alpha_{1,2}(1) = 0.7855.$$

Using these hyper parameters and the data, we find that the posterior law is a $\text{Beta}(18.2417, 3.7855)$. This is very close to the non-informative estimation $\text{Beta}(18, 4)$ given in Table 2. Hence, even if the sample size data seems limited,there is enough information to limit the sensitivity of the prior. Note that we can nevertheless force $(\alpha_{i,j})$ to be large enough to have an impact.

We then restrict ourselves to the non-informative prior because it gives similar results and several incomes are unknown as for instance for choosing $(\alpha_{1,5})$.

We can now see the evolution of the population over short time spans. To illustrate this, let us represent for each year t, the prediction we give from the previous years $2005, \dots, t - 1$. We represent these successive predictions in Figure 3. Each boxplot represents the distribution of the prediction. The mean estimator given by Formula (10) was added (diamond symbol). To compare with the genuine evolution of the population, we add the total number of female bears at each year (red circles). In this figure, we can for instance see the learning step of our algorithm.

Let us now focus on λ; that is the viability problem. For this model the principal eigenvalue λ is then the solution of the equation

$$-\lambda^5 + \lambda^4 p_{5,5}(1) + \sum_{k=1}^{4} k p_{5,1}(k)) p_{1,2}(1) p_{2,3}(1) p_{3,4}(1) p_{4,5}(1) = 0,$$

which has nevertheless no explicit solution (this equation comes from the fact that λ is just the root of the characteristic polynomial). However, numerically, using the algorithm given in Appendix 8.2 one can find

$$\mathbb{P}(\lambda \leq 1) = 0.012.$$

Hence, with high probability the process is super-critical and the reproduction capacity seems to guarantee viability. However if the population is too small, the survival is not guaranteed. Again, surviving probabilities s in Equation (7) have no explicit solutions but are functions of $p_{i,j}$ and one can use the algorithm of Appendix 8.2 to find that, with the current population size, the probability of extinction can be rounded to 0. It seems that the size of the population and the reproduction capacity is large enough to guaranty survival without reintroduction in the central part of the Pyrenees. Let us emphasize again that this result is only valid under our main assumptions. That includes no change in the environmental conditions including hunting or other human interference. Note that we also assumed no change in the basic demographic parameters that could be due to inbreeding depression. In fact, the current population is highly inbreed [10].

However one can still see in Figure 4 that it is more beneficial to reintroduce older bears than younger ones. In particular, in 2018, two females were reintroduced in the western part of the Pyrenees. Assuming similar evolution, there is around 15% risk of extinction for this sub-population (of this two males and 2 reintroduced females). Additionally, if these females are pregnant and then if they survive and give birth each to one female cub then this risk will be divided by 2 (around 8%). Closely related to theses questions, using Formula (7), we can calculate the effective population size; namely the number of reproducing individuals we need to have an extinction risk lower than a defined threshold. Here, to have an extinction probability lower than 0.05 then we need at least 5 reproducing females.

6. **Extension to different types of dataset** To apply the present method, we need to know exactly the sum over the time of the number of individuals passing from type i to j at each time. This type of information is almost as difficult to collect as the complete genealogical tree of each individual. Even if several such datasets exist and to the best of our knowledge, few or no previous methods exhaustively use complete information. It is not always possible to have access to this type of full dataset [10, 14, 21]. We explain some extensions to generalize our approach in this section. This was not done simply to clarify.

In the synthetic example of Section 5, we have to know about the whole life of each individual (survival and reproduction) but in general we only know the number of survivors and the number of offspring, then we can write

$$p_{1,1} = p_S + p_R,$$

as seen written in this example. However, suppose that p_S and p_R are both Dirichlet distributed and independent, this is not the same modelling framework and with this modelling choice, we can study survival and reproduction separately.

Another generalisation comes from the second example of the French Pyrenean brown bears. In general we do not know the sex of the bear cub before until a few years after their birth. Consequently, we do not know $n_{4,1}(\cdot, t)$ for the last year, t. However, one can slightly change the Bayesian model to take into account this missing information. Indeed, one can add some structure and assume that

$$p_{4,1}(k) = \sum_{l \geq k} q(l) \binom{l}{k} p^k (1-p)^{l-k},$$

where q is the law of the number of spring for a female bear (male and female bear cubs) and p is the probability that a cub is a female. Using this (natural) representation then we keep the conjugation property, and all the posterior remains explicit. For the dataset [10], this leads, for instance, to 0.583 as posterior mean estimator for the sex ratio. This is slightly more favourable than the estimation of 0.5 used in [15]. Using the number of males in the population and supposing that the survival rates are similar can reduce the variance of our estimation; this is a compromise between the data and expertise.

The assumption that we have made, which can be restrictive, is that surviving and reproducing are independent, but for prediction in contrast with parameter interpretation, this is not a real problem. Nevertheless, one can extend our setting to model the necessity of survival to make reproduction possible. The assumption that all $p_{i,j}$ are independent from each other can be removed. A careful reading of the proof of Formula (9) shows that we can extend it for the following setting: for every $k_1, ..., k_K$, we replace all $p_{i,j}(k)$ by $p_i(k_1, ..., k_K)$ which is the probability that an individual with type i has k_1 offspring of type 1 and k_2 offspring of type 2 $etc.$ It remains to assume that all p_i are independent and keep a Dirichlet law (on $\prod_{j=1}^{K} \{1, .., \kappa_{i,j}\}$ instead of $\{1, ..., \kappa_{i,j}\}$) as prior to have the same type of approach.

Finally, In many examples coming from ecology, the number of offspring may be very large and it is not possible to fix any value $\kappa_{i,j}$ for the maximum numbers of offspring. It is therefore not possible to directly use our approach unless we take a large $\kappa_{i,j}$ (this requires too much information). In order to keep the same idea for our approach, we can suppose that offspring are

Times t	0	1	2	3	4	5
N	100	75	59	43	33	22
$n_{1,1}(0,t)$	47	30	29	20	18	
$n_{1,1}(1,t)$	39	37	23	17	11	
$n_{1,1}(2,t)$	8	4	2	3	2	
$n_{1,1}(3,t)$	4	2	4	2	1	
$n_{1,1}(4,t)$	2	2	1	1	1	

TABLE 1. Synthetic data of Section 5, $N(t)$ represents the number of individuals at time t and $\mathbf{n} = (n_{1,1}(k,t))_{0\leq k\leq 4, 0\leq t\leq 4}$ the learning data.

given by a Poisson law $\mathcal{P}(m)$ (which arises naturally as $\kappa_{i,j} \to \infty$ as a limit of our model under general assumptions [3]). Also the naturally associated prior will be the Gamma law. It also verifies properties (conjugation, non-informative, expertise, $etc.$) developed in Section 3 and we also have an explicit posterior distribution.

7. **Discussion** We have shown how the Galton-Watson process with Bayesian inference is a simple but powerful setting for PVA analysis. On top of giving the same information as the matrix population model, it easily gives precise results for probability of extinction, time to extinction, $etc.$ It is perfectly adapted for small populations and then perfectly completes previous studies, see [9, 19, 24, 28, 39]. Deterministic models including differential equation and matrix products are well suited to large population (but sometimes more than 10^6 individuals [11]). Stochastic differential equations or hybrid models (including piecewise deterministic models) are more appropriate at a mesoscopic scale; at least for a thousand individuals. Few models are concerned with smaller population size although these populations are more concerned by extinction.

For conservation purposes, there is more and more population monitoring and there is more and more exhaustive information to understand population evolution. Nevertheless, classical statistical methods do not use this exhaustive information to calibrate demographic model. The Bayesian setting we present offers the advantage of taking into account all the information of these types of dataset and can be adapted for missing data or less informative data.

The combination between Galton-Watson processes and Bayesian, and known mathematical results [2, 26, 27, 36] implies simple formulas and there is no need for large simulations to answer simple questions. Results are found without using long computational algorithms with unknown errors. Consequently, it is easily scalable to big data: in contrast with ABC methods [20] or other MCMC methods [44], there is no need for a large number of simulations; this permits us to use this method for large datasets. One can easily use this model as part of a larger model to motivate control of the population or others strategies. Furthermore, there is no doubt about the convergence of this type of algorithm and the associated errors due to such computations.

For small data sets, our approach enables to add expert knowledge and gives precise confidence intervals or error estimation for the prediction. This takes into account statistical errors and the demographic stochasticity. However, the environmental stochasticity is not directly included in this estimation even if the Bayesian learning picks up some part of it with the data.

This model is founded on several assumptions and in general, some of them can be questionable such as the absence of carrying capacity, the independence of individuals and the static environment but they are very natural considering a small population during short time period. However, genetic aspects that can be of importance for small populations is not taken into account. Considering the example of the bears, we do not speak about the male reintroduction which can have a real interest to avoid inbreeding depression. We focus on demographic aspects and numbers. Including the inbreeding issue in our setting may be feasible and could provide interesting insights [16, 27].

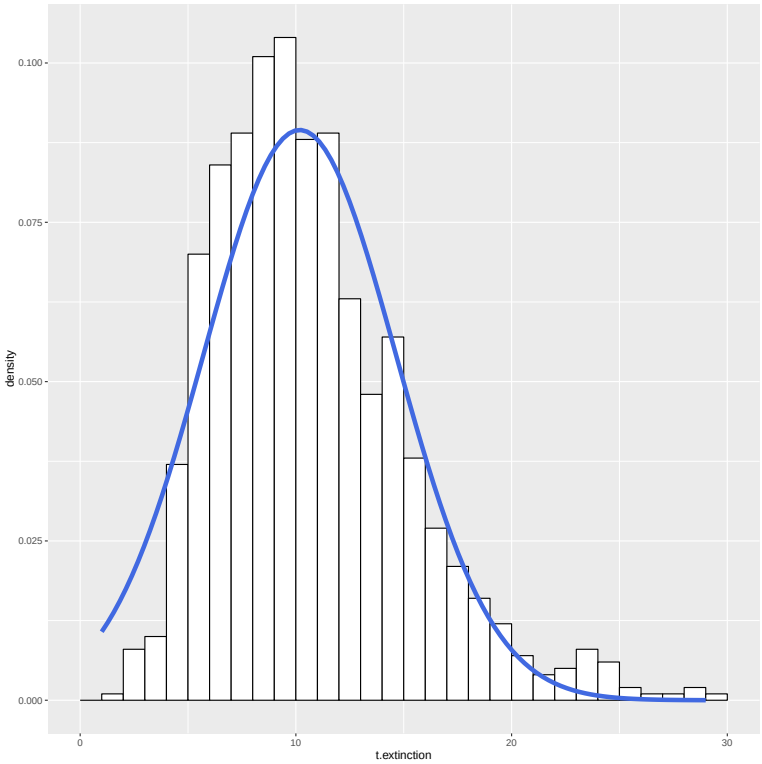


FIGURE 1. The histogram represents the law of the extinction time for the simulated example. The blue curve is the normal distribution with same mean and variance.

Parameter	Posterior	Mean	IC 90%
$p_{0,1}(1)$	Beta (18,4)	0.8181	[0.67,0.93]
$p_{1,2}(1)$	Beta(17,1)	0.9444	[0.84,1]
$p_{2,3}(1)$	Beta(13,3)	0.8125	[0.64,0.94]
$p_{3,4}(1)$	Beta(13,1)	0.9286	[0.79,1]
$p_{4,4}(1)$	Beta(73,4)	0.9480	[0.90,0.98]
$p_{4,0}$	Dir(71,9,7,1)		
$p_{4,0}(0)$	Beta(71,17)	0.8068	[0.73,0.87]
$p_{4,0}(1)$	Beta(9,79)	0.1023	[0.05,0.16]
$p_{4,0}(2)$	Beta(7,81)	0.0795	[0.04,0.13]
$p_{4,0}(3)$	Beta(1,87)	0.0114	[0,0.03]

TABLE 2. Posterior for all parameters $p_{i,j}(k)$ and the vector valued $p_{4,0}$ for the bears population example. We use the data [10] and the non-informative prior.

8. **Appendix**

8.1. *Proof of the expression of the Dirichlet distribution*

First we recall that

$$p_{i,j} \sim \mathbf{Dir}\left(\alpha_{i,j}(0), ..., \alpha_{i,j}(\kappa_{i,j})\right),$$

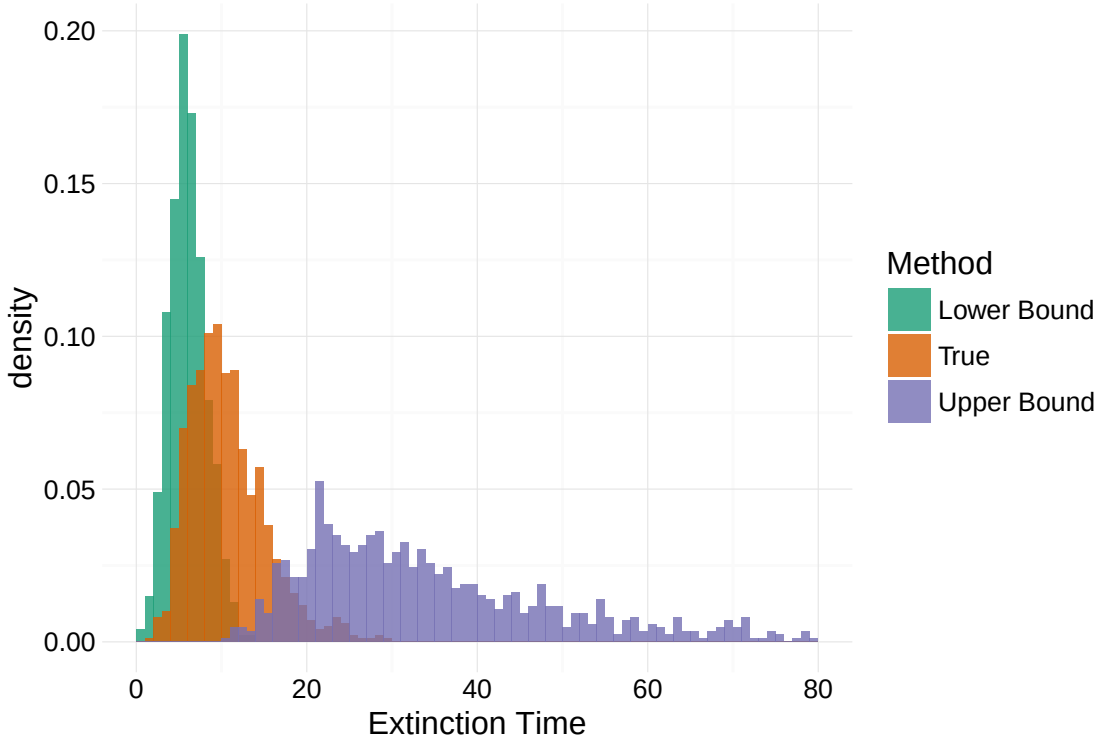


FIGURE 2. The orange histogram represents the law of the extinction time. The other two histograms represent the law of our lower and upper bound. One can see that they intertwine with the beginning and the end of the middle histogram.

means that for all $a_1, ...a_{\kappa_{i,j}}, b_1, ..., b_{\kappa_{i,j}}$, setting $A = [a_1, b_1] \times \cdots \times [a_{\kappa_{i,j}}, b_{\kappa_{i,j}}]$, we have

$$\begin{aligned}&\mathbb{P}(p_{i,j} \in A)\\ &= \int_A \frac{\Gamma(\sum_{k=1}^{\kappa_{i,j}} \alpha_{i,j}(k))}{\prod_{k=1}^{\kappa_{i,j}} \Gamma(\alpha_{i,j}(k))}\\ &\quad \times \prod_{k=1}^{\kappa_{i,j}} x_k^{\alpha_{i,j}(k)-1} dx_1 \dots dx_{\kappa_{i,j}},\end{aligned}$$

where Γ is the classical Gamma function. Then, by definition of the model we have the following prior:

$$\mathcal{L}(p_{i,j}) = \mathbf{Dir}\left(\alpha_{i,j}(0), ..., \alpha_{i,j}(\kappa_{i,j})\right).$$

Let us now detail the likelihood of the model using multinomial laws. Let us denote by $N_{i,j}(t,k)$ the number of individuals with type i having j descendants of type i at time t. Namely

$$N_{i,j}(t,k) = \sum_{l=1}^{N_i(t)} \mathbf{1}_{\{\xi_{i,j,l,t}=k\}}.$$

We have for all $t \geq 0$

$$\begin{aligned}&\mathcal{L}\left((N_{i,j}(k,t))_{0\leq k\leq \kappa_{i,j}} \mid (p_{i,j}(k))_{0\leq k\leq \kappa_{i,j}}, N_i(t)\right)\\ =&\mathbf{M}(N_i(t), (p_{i,j}(k))_{0\leq k\leq \kappa_{i,j}}),\end{aligned}$$

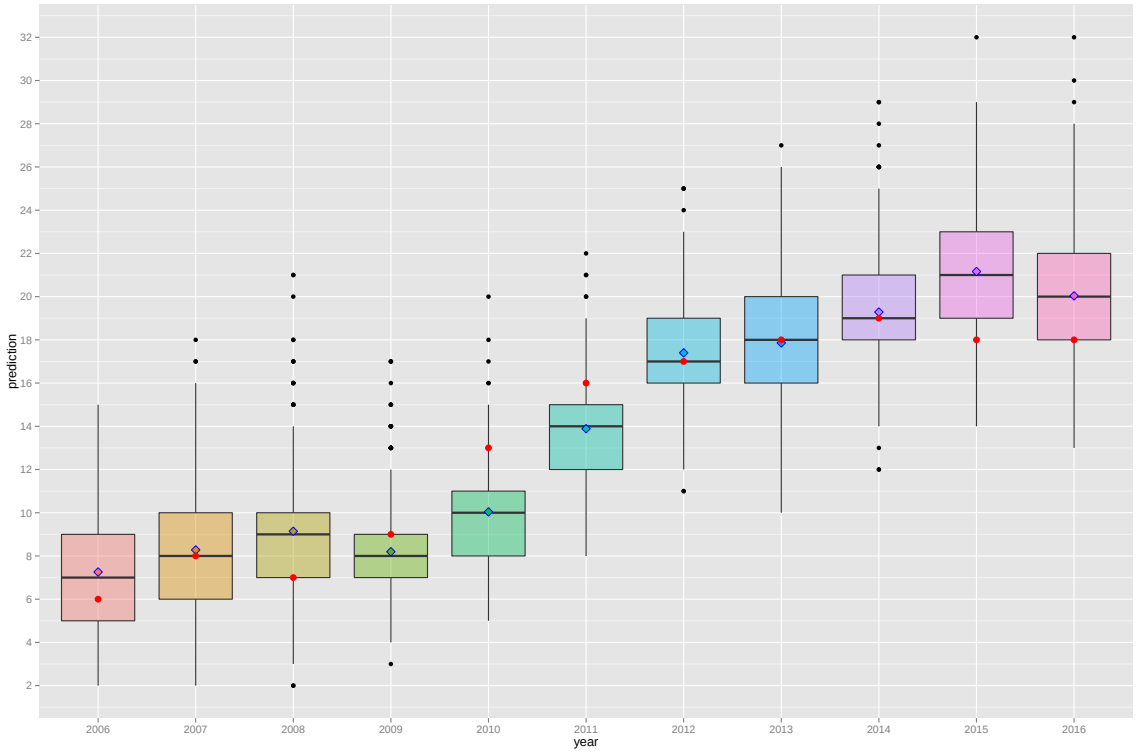


FIGURE 3. Red circles represent the total number of female bears at each year. Boxplots and diamonds both represent prediction of the number pf females at each year considering data available on the previous years and a non-informative prior. The boxplot represents the law of the prediction and diamond represents the mean estimator given by Formula (10).

where $\mathbf{M}(n, q_0, ..., q_\kappa)$ denotes the usual multinomial law with n trials and $q_0, ..., q_m$ events probabilities. Namely $M \sim \mathbf{M}(n, q_0, ..., q_\kappa)$ means that for all $m_0, \dots, m_\kappa$ we have

$$\begin{aligned}&\mathbb{P}\left(M(0) = m_0, \dots, M(\kappa) = m_\kappa\right)\\ =&\frac{n!}{m_0! \dots m_\kappa!} q_0^{m_0} \times \dots \times q_\kappa^{m_\kappa},\end{aligned}$$

when $\sum_{i=0}^{\kappa} m_i = n$ either it is equal to 0. Using these two explicit formulas (multinomial and Dirichlet densities) and the Bayes Theorem, it is easy to check the classical conjugation result:

$$\begin{aligned}&\mathcal{L}\left((p_{i,j}(k))_{1\leq k\leq \kappa_{i,j}}, \mid (N_{i,j}(k,t))_{1\leq k\leq \kappa_{i,j}}, N_i(t)\right)\\ =&\mathbf{Dir}\left(\alpha_{i,j}(0) + N_{i,j}(0,t), ..., \alpha_{i,j}(\kappa_{i,j}) + N_{i,j}(\kappa_{i,j}, t)\right).\end{aligned}$$

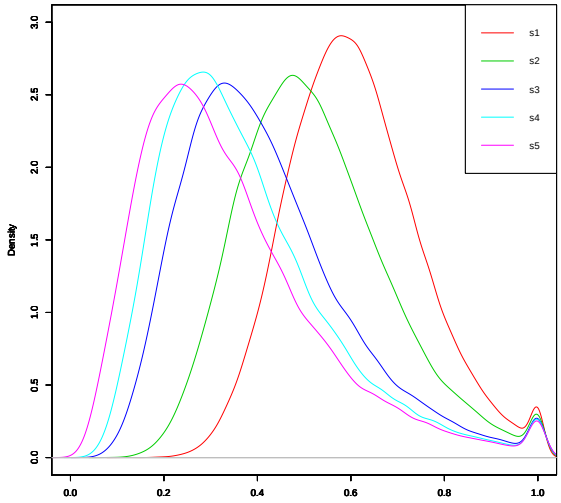


FIGURE 4. Density of parameter $s_1, ..., s_5$ for the bear example of Section 5. Data comes from [10] and we use a non-informative prior.

Further details can be found in [41, 42]. It remains to prove that it holds true for any times through a chain rule type argument. The Markov property (derived by the independence assumption) gives that

$$\begin{aligned}
&\mathcal{L}\left((N_{i,j}(k,t))_{0\le k\le \kappa_{i,j},0\le t\le T} \mid (p_{i,j}(k))_{0\le k\le \kappa_{i,j}}\right)\\
=&\mathcal{L}\left((N_{i,j}(k,t))_{0\le k\le \kappa_{i,j},0\le t\le T} \mid (N_{i,j}(k,t))_{0\le k\le \kappa_{i,j}}, (p_{i,j}(k))_{0\le k\le \kappa_{i,j}}\right)\\
\otimes&\,\mathcal{L}\left((N_{i,j}(k,t))_{0\le k\le \kappa_{i,j},0\le t\le T-1} \mid (p_{i,j}(k))_{0\le k\le \kappa_{i,j}}\right)\\
=&\mathcal{L}\left((N_{i,j}(k,T))_{0\le k\le \kappa_{i,j}} \mid N_i(T), (p_{i,j}(k))_{0\le k\le \kappa_{i,j}}\right)\\
\otimes&\,\mathcal{L}\left((N_{i,j}(k,t))_{0\le k\le \kappa_{i,j},0\le t\le T-1} \mid (p_{i,j}(k))_{0\le k\le \kappa_{i,j}}\right)\\
=&\prod_{t=1}^{T}\mathcal{L}\left((N_{i,j}(k,t))_{0\le k\le \kappa_{i,j},0} \mid N_i(t), (p_{i,j}(k))_{0\le k\le \kappa_{i,j}}\right).
\end{aligned}$$

Using this formula, the expressions of multinomial and Dirichlet densities and Bayes Theorem again finalise the proof of Equation (9). Note that $\mathbf{n}$ that represents the data is a realisation of the random variables $(N_{i,j}(k,t))_{0\le k\le \kappa_{i,j},0\le t\le T}$.

8.2. *Pathways to PVA through the Bayesian-Galton-Watson approach: the expanded version*

Here we give details for each step of Box 1.

Previous knowledge

Initializes $(\alpha_{i,j}(k))$ as follows:

i) If one cannot pass from state i to state j, set

$$p_{i,j}(0) = 1.$$

or equivalently, in a certain sense,

$$\alpha_{i,j}(k) = \mathbf{1}_{k=0}.$$

ii) If you have no information, set

$$\alpha_{i,j}(k) = 1.$$

iii) If we use previous estimates and we expect that $p_{i,j} = m_{i,j}(k) \pm \epsilon_{i,j}(k)$ and $\epsilon_{i,j}(k)$ is centred and has variance of magnitude

$$\sigma_{i,j}^2 m_{i,j}(k)(1 - m_{i,j}(k))$$

then set

$$\alpha_{i,j}(k) = \frac{(1 - \sigma_{i,j}^2) m_{i,j}(k)}{\sigma_{i,j}^2},$$

if $\sigma_{i,j}^2 < 1$ else $\alpha_{i,j}(k) = 1$.

iv) If we want to impose an expert opinion. The experts think that

$$p_{i,j} \approx q_{i,j}$$

and that their opinion is as important as M data then we can choose

$$\alpha_{i,j} = M \times q_{i,j}.$$

We can see in case iii), that this choice enables to match the expectation to the variance. More precisely, to match the expectation, we need

$$\frac{\alpha_{i,j}(k)}{S_{i,j}} = m_{i,j}(k)$$

with $S_{i,j} = \sum_{k=0}^{\kappa_{i,j}} \alpha_{i,j}(k)$. We have then $\alpha_{i,j}(k) = S_{i,j} m_{i,j}(k)$. To match the variance, we have

$$\begin{aligned} &\frac{S_{i,j} m_{i,j}(k)(S_{i,j} - S_{i,j} m_{i,j}(k)}{S_{i,j}^2 (S_{i,j} + 1)} \\ =&m_{i,j}(k)(1 - m_{i,j}(k))\sigma_{i,j}^2, \end{aligned}$$

hence

$$S_{i,j} = \frac{1}{\sigma_{i,j}^2} - 1 = \frac{1 - \sigma_{i,j}^2}{\sigma_{i,j}^2}.$$

Of course we need $\sigma_{i,j}^2 < 1$. However, if this is not the case, the noise is too large and the information on $p_{i,j}$ is therefore less informative than no information.

Calculus of quantities of interest

For K ($K = 1$ or 2), one can directly give an expression for almost all quantities through some integral calculus. However, in general, it is difficult to express them. The better way to calculate this quantity is perhaps the following simple Monte Carlo algorithms.

Due to the central limit theorem the error of this simple algorithm can be estimated through an estimator of the variance but in general one cannot fix a number of simulations before doing the calculation. However for the probability of viability and the probability of extinction, the variance can be bounded by $1/4$ and then the error of this simple algorithm is lower than $(4\sqrt{n_{\text{prec}}})^{-1}$. For an error lower than 0.5% we can take $n_{\text{prec}} = 2500$.

Below, simulate $p_{i,j}$ means drawing a random variable with law

$$\mathbf{Dir}\left(\alpha_{i,j}(0) + \sum_{t=0}^{T} n_{i,j}(0,t), \ldots, \alpha_{i,j}(\kappa_{i,j}) + \sum_{t=0}^{T} n_{i,j}(\kappa_{i,j}, t)\right).$$

Short time evolution:

Do n_{prec} times the steps

- Simulate the K^2 probability valued random variables $p_{i,j}$.
- Calculate $M^n \cdot X_0$ with M defined by (4).

Then do a mean value of these proportions.

Probability of viability

Do n_{prec} times the steps

- Simulate the K^2 probability valued random variables $p_{i,j}$.
- Take for λ the largest eigenvalue of the M defined by (4).

Then count the proportion of λ being larger than 1.

Time to extinction: To bound the extinction time T_{Ext} between two values

$$T_- \leq T_{\text{Ext}} \leq T_+,$$

with probability $1-\alpha$, for some fixed threshold α, do n_{prec} times the steps

- Simulate the K^2 probability valued random variables $p_{i,j}$.
- Calculate the function

$$U(t) = \lambda^t \frac{\sum_{j=1}^K N_j(T) v_j}{\min(v)}$$

and

$$L(t) = \frac{\max(v)^2}{\min(v)^2} \frac{(1-\lambda)}{\Xi} \lambda^{t+1} \sum_{j=1}^K N_j(T) v_j,$$

where

$$\Xi = \sum_{j=1}^K \frac{v_j^2}{\min(v)} \sup_{1\leq i\leq K} \sum_{k=1}^{\kappa_{i,j}} (k^2 - M_{i,j}^2) p_{i,j}(k)$$

where v is given in Equation (5).

Then do a mean value of functions U and L and choose (T_-, T_+) such that

$$L(T_-) = \alpha, \qquad U(T_+) = \alpha.$$

These two functions are based on Equation (6) and Lemma 8.1. Note that these bounds on extinction time only hold when $\lambda < 1$. To avoid some computational problems due to the fact that $\mathbb{P}(\lambda < 1) > 0$ (even if it is very small), it can be useful, when calculating the extinction time conditioned on the event $\{\lambda < 1\}$. In the previous algorithm, this conditioning translates into doing a mean value only for λ satisfying $\lambda < 1$.

Probability of extinction:

Do n_{prec} times the steps

- Simulate the K^2 probability valued random variables $p_{i,j}$.
- Find the (vector) solution s of equation $\varphi(s) = s$, with φ is the generating function defined in (8). To do this step one can use any classical optimization algorithm.
- Calculate $s_1^{N_1(T)} \times \cdots \times s_K^{N_K(T)}$.

Then do a mean of these quantities.

How to plan reintroduction: Do n_{prec} times the steps

- Simulate the K^2 probability valued random variables $p_{i,j}$.
- Find the (vector) solution s of equation $\varphi(s) = s$, with φ is the generating function defined in (8). To do this step one can use any classical optimization algorithm.

Then do a histogram or smoothed density through classical kernel density methods for instance.

8.3. *Lower bound for the extinction time*

Lemma 8.1. *We have*

$$\mathbb{P}\left(T_{Ext} \geq t\right) \geq \frac{\max(v)^2}{\min(v)^2} \frac{\lambda^{2t} \left(\sum_{j=1}^K v_j N_j(0)\right)^2}{\Xi \lambda^{t-1} \left(\frac{1-\lambda^t}{1-\lambda}\right) \left(\sum_{j=1}^K v_j N_j(0)\right) + \lambda^{2t} \left(\sum_{j=1}^K v_j N_j(0)\right)^2}, \tag{11}$$

where

$$\Xi = \sum_{j=1}^K \frac{v_j^2}{\min(v)} \sup_{1\leq i\leq K} \sum_{k=1}^{\kappa_{i,j}} (k^2 - M_{i,j}^2) p_{i,j}(k).$$

Moreover, the right-hand side of equation (11) *is equivalent to*

$$\frac{\max(v)^2}{\min(v)^2} \frac{\lambda^{2t} \left(\sum_{j=1}^K v_j N_j(0)\right)^2}{\Xi \lambda^{t-1} \left(\frac{1-\lambda^t}{1-\lambda}\right) \left(\sum_{j=1}^K v_j N_j(0)\right) + \lambda^{2t} \left(\sum_{j=1}^K v_j N_j(0)\right)^2} \underset{t\to\infty}{\sim} \frac{\max(v)^2}{\min(v)^2} \lambda^{t+1} \left(\sum_{j=1}^K v_j N_j(0)\right) \frac{(1-\lambda)}{\Xi}.$$

Proof. The proof is inspired from [26, Box 5.2 p.119] We have

$$\mathbb{P}\left(\sum_{j=1}^{K} N_j(t) \geq 1\right) \geq \frac{\max(v)^2}{\min(v)^2} \frac{\mathbb{E}[X(t)]^2}{\mathbb{E}[X(t)^2]},$$

where

$$X(t) = \sum_{j=1}^{K} v_j N_j(t).$$

Yet, we have $\mathbb{E}[X(t)] = \lambda^t \sum_{j=1}^{K} v_j N_j(0)$, and it remains to study the second moment. Recall

$$N_j(t+1) = \sum_{i=1}^{K} \sum_{l=1}^{N_i(t)} \xi_{i,j,l,t}.$$

Setting $\mathbf{N}(t) = \sum_j N_j(t)$ for the population size and reindexing the population, we find

$$N_j(t+1) = \sum_{l=1}^{\mathbf{N}(t)} \zeta_{i(l),j,l,t},$$

where $i(l)$ is the type of the individual l and ζ its offspring. We have

$$\begin{aligned}
\mathbb{E}[X(t+1)^2] &= \mathbb{E}\left[\left(\sum_{l=1}^{\mathbf{N}(t)} \sum_{j=1}^{K} v_j \zeta_{i(l),j,l,t}\right)^2\right] \\
&= \mathbb{E}\left[\sum_{l,l'=1}^{\mathbf{N}(t)} \sum_{j,j'=1}^{K} \left(v_j \zeta_{i(l),j,l,t}\right)\left(v_{j'} \zeta_{i(l'),j',l',t}\right)\right] \\
&= \mathbb{E}\left[\sum_{l,l'=1}^{\mathbf{N}(t)} \sum_{j,j'=1}^{K} \mathbb{E}\left[\left(v_j \zeta_{i(l),j,l,t}\right)\left(v_{j'} \zeta_{i(l'),j',l',t}\right]\right])\right] \\
&= \mathbb{E}\left[\sum_{l,l'=1}^{\mathbf{N}(t)} \sum_{j,j'=1}^{K} v_j v_{j'} (\mathbf{1}_{l=l'} M^{(2)}_{i(l),j,j'} + \mathbf{1}_{l\neq l'} M_{i(l),j} \times M_{i(l'),j'})\right] \\
&= \mathbb{E}\left[\sum_{j,j'=1}^{K} v_j v_{j'} \left(\sum_{i=1}^{K} N_i(t) M^{(2)}_{i,j,j'} + \sum_{i,i'=1}^{K} N_i(t) N_{i'}(t) M_{i,j} \times M_{i',j'} - \sum_{i=1}^{K} N_i(t) M_{i,j} \times M_{i',j'}\right)\right].
\end{aligned}$$

In the third line we conditioned on $\mathbf{N}(t)$ and use the independance property. In the fourth line, we use the notation

$$M^{(2)}_{i(l),j,j'} = \mathbb{E}\left[\left(v_j \zeta_{i(l),j,l,t}\right)\left(v_{j'} \zeta_{i(l'),j',l',t}\right)\right],$$

and then under our assumptions

$$M^{(2)}_{i(l),j,j'} = M_{i(l),j} M_{i(l),j'},$$

if $j \neq j'$ and either

$$M^{(2)}_{i(l),j,j} = \sum_{k=0}^{\kappa_{i(l),j}} k^2 p_{i,j}(k).$$

Then

$$\begin{aligned}
&\sum_{j,j'=1}^{K} v_j v_{j'} \left(\sum_{i=1}^{K} N_i(t) M^{(2)}_{i,j,j'} + \sum_{i,i'=1}^{K} N_i(t) N_{i'}(t) M_{i,j} \times M_{i',j'} - \sum_{i=1}^{K} N_i(t)^2 M_{i,j} \times M_{i',j'} \right) \\
=&\lambda^2 X(t)^2 + \sum_{j,j'=1}^{K} v_j v_{j'} \left(\sum_{i=1}^{K} N_i(t) M^{(2)}_{i,j,j'} - \sum_{i=1}^{K} N_i(t) M_{i,j} \times M_{i',j'} \right) \\
=&\lambda^2 X(t)^2 + \sum_{j=1}^{K} \sum_{i=1}^{K} N_i(t) v_j^2 \left(M^{(2)}_{i,j,j'} - M_{i,j} \times M_{i',j'} \right) \\
=&\lambda^2 X(t)^2 + \sum_{j=1}^{K} \sum_{i=1}^{K} N_i(t) v_j^2 \sigma^2_{i,j}.
\end{aligned}$$

and also setting $\overline{\sigma_j^2} = \sup_i \sigma^2_{i,j}$

$$\sum_{j=1}^{K} \sum_{i=1}^{K} N_i(t) v_j^2 \sigma^2_{i,j} \leq \sum_{i=1}^{K} N_i(t) v_i \sum_{j=1}^{K} v_j^2 \overline{\sigma_j^2} / \min(v).$$

Finally setting $\Xi = \sum_{j=1}^{K} v_j^2 \overline{\sigma_j^2} / \min(v)$, we obtain

$$\begin{aligned}
\mathbb{E}[X(t+1)^2] &\leq \lambda^2 \mathbb{E}[X(t)^2] + \Xi \mathbb{E}[X(t)] \\
&\leq \lambda^2 \mathbb{E}[X(t)^2] + \Xi \lambda^t X(0) \\
&\leq \Xi \lambda^t \left(\frac{1-\lambda^{t+1}}{1-\lambda} \right) X(0) + \lambda^{2(t+1)} X(0)^2
\end{aligned}$$

where we used an iteration argument in the third line. To conclude

$$\begin{aligned}
\mathbb{P}\left(\sum_{j=1}^{K} N_j(t) \geq 1 \right) &\geq \frac{\max(v)^2}{\min(v)^2} \frac{\lambda^{2t} X(0)^2}{\Xi \lambda^{t-1} \left(\frac{1-\lambda^t}{1-\lambda} \right) X(0) + \lambda^{2t} X(0)^2} \\
&\sim \frac{\max(v)^2}{\min(v)^2} \lambda^{t+1} X(0) \frac{(1-\lambda)}{\Xi}.
\end{aligned}$$

□

References

[1] S. Asmussen and N. Keiding. Martingale central limit theorems and asymptotic estimation theory for multitype branching processes. *Advances in Applied Probability*, 10(1):109–129, 1978.

[2] K. B. Athreya, P. E. Ney, and P. Ney. *Branching processes*. Courier Corporation, 2004.

[3] A. Barbour, L. Holst, and S. Janson. *Poisson Approximation*. Oxford science publications. Clarendon Press, 1992.

[4] H. Bernardelli. Population waves. *J. Burma Res. Soc.*, 31:3–18, 1941.

[5] I.-J. Bienaymé. De la loi de multiplication et de la durée des familles. *Soc. Philomat. Paris Extraits, Sér*, 5:37–39, 1845.

[6] P. Bliman, M. S. Aronna, F. C. Coelho, and M. A. H. B. da Silva. Global stabilizing feedback law for a problem of biological control of mosquito-borne diseases. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 3206–3211, Dec 2015.

[7] M. S. Boyce. Restitution of gamma-and k-selection as a model of density-dependent natural selection. *Annual review of ecology and systematics*, 15(1):427–447, 1984.

[8] M. S. Boyce. Population viability analysis. *Annual Review of Ecology and Systematics*, 23(1):481–497, 1992.

[9] B. Cairns, J. Ross, and T. Taimre. A comparison of models for predicting population persistence. *Ecological Modelling*, 201(1):19 – 26, 2007. Management, Control and Decision Making for Ecological Systems.

[10] J. CAMARRA, J. SENTILLES, A.GASTINEAU, and P. QUENETTE. Suivi de l'ours brun dans les pyrenees francaises (sous-populations occidentale et centrale) rapport annuel année 2016. Technical report, O.N.C.F.S. - Unité Prédateurs et Animaux Déprédateurs - Equipe Ours, 2016.

[11] F. Campillo and C. Lobry. Effect of population size in a predator–prey model. *Ecological Modelling*, 246:1 – 10, 2012.

[12] M. L. Carvalho. A joint estimator for the eigenvalues of the reproduction mean matrix of a multitype galton-watson process. *Linear algebra and its applications*, 264:189–203, 1997.

[13] H. Caswell. *Matrix population models: construction, analysis, and interpretation*. Sinauer Associates, 1989.

[14] H. Caswell, M. Fujiwara, and S. Brault. Declining survival probability threatens the north atlantic right whale. *Proceedings of the National Academy of Sciences*, 96(6):3308–3313, 1999.

[15] G. Chapron, P.-Y. Quenette, S. Legendre, and J. Clobert. Which future for the french pyrenean brown bear (ursus arctos) population? an approach using stage-structured deterministic and stochastic models. *Comptes Rendus Biologies*, 326:174 – 182, 2003.

[16] D. Daley. Extinction conditions for certain bisexual galton-watson branching processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 9(4):315–322, 1968.
[17] D. DeMaster, R. Angliss, J. Cochrane, P. Mace, R. Merrick, M. Miller, S. Rumsey, B. Taylor, G. Thompson, and R. Waples. Recommendations to noaa fisheries: Esa listing criteria by the quantitative working group 10 june 2004. Technical report, Recommendations to NOAA Fisheries: ESA Listing Criteria by the Quantitative Working Group, 10 June 2004. U.S. Dep. Commerce, NOAA Tech. Memo. NMFSF/SPO-67, 2004.
[18] J. Dion. Statistical inference for discrete time branching processes. *Lecture Notes of the Seventh International Summer School on Probability Theory and Mathematical Statistics. Singapore: Sci. Culture Technology Publishing*, pages 60–121, 1993.
[19] S. Engen, R. Lande, B.-E. Sæther, and H. Weimerskirch. Extinction in relation to demographic and environmental stochasticity in age-structured models. *Mathematical Biosciences*, 195(2):210–227, 2005.
[20] A. ESTOUP, E. LOMBAERT, J.-M. MARIN, T. GUILLEMAUD, P. PUDLO, C. P. ROBERT, and J.-M. CORNUET. Estimation of demo-genetic model probabilities with approximate bayesian computation using linear discriminant analysis on summary statistics. *Molecular Ecology Resources*, 12(5):846–855.
[21] C. A. H. Foley and L. J. Faust. Rapid population growth in an elephant loxodonta africana population recovering from poaching in tarangire national park, tanzania. *Oryx*, 44(2):205–212, 2010.
[22] P. Foley. Predicting extinction times from environmental stochasticity and carrying capacity. *Conservation Biology*, 8(1):124–137, 1994.
[23] E. Gerber, C. Krebs, C. Murrell, M. Moretti, R. Rocklin, and U. Schaffner. Exotic invasive knotweeds (fallopia spp.) negatively affect native plant and invertebrate assemblages in european riparian habitats. *Biological Conservation*, 141(3):646–654, 2008.
[24] L. R. Ginzburg, L. B. Slobodkin, K. Johnson, and A. G. Bindman. Quasiextinction probabilities as a measure of impact on population growth. *Risk Analysis*, 2(3):171–181.
[25] F. Gosselin and J.-D. Lebreton. Potential of branching processes as a modeling tool for conservation biology. In *Quantitative methods for conservation biology*, pages 199–225. Springer, 2000.
[26] P. Haccou, P. Haccou, P. Jagers, V. A. Vatutin, and V. A. Vatutin. *Branching processes: variation, growth, and extinction of populations*. Number 5. Cambridge university press, 2005.
[27] T. E. Harris. *The theory of branching processes*. Courier Corporation, 2002.
[28] E. E. Holmes, J. L. Sabo, S. V. Viscido, and W. F. Fagan. A statistical approach to quasi-extinction forecasting. *Ecology Letters*, 10(12):1182–1198.
[29] M. Kimmel and D. E. Axelrod. Branching processes in biology (interdisciplinary applied mathematics). 2002.
[30] J.-D. Lebreton, F. Gosselin, and C. Niel. Extinction and viability of populations: paradigms and concepts of extinction models. *Ecoscience*, 14(4):472–481, 2007.
[31] P. H. Leslie. On the use of matrices in certain population mathematics. *Biometrika*, pages 183–212, 1945.
[32] E. G. Lewis. On the generation and growth of a population. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 6(1):93–96, 1942.
[33] S. Lowe, M. Browne, S. Boudjelas, and M. De Poorter. *100 of the world's worst invasive alien species: a selection from the global invasive species database*, volume 12. Invasive Species Specialist Group Auckland, 2000.
[34] F. Maaouia and A. Touati. Identification of multitype branching processes. *Ann. Statist.*, 33(6):2655–2694, 12 2005.
[35] G. M. MACE and R. LANDE. Assessing extinction threats: Toward a reevaluation of iucn threatened species categories. *Conservation Biology*, 5(2):148–157.
[36] S. Méléard. *Modèles aléatoires en Ecologie et Evolution*. Springer, 2016.
[37] M. Mendoza and E. Gutiérrez-Pena. Bayesian conjugate analysis of the galton-watson process. *Test*, 9(1):149–171, 2000.
[38] W. F. Morris, D. F. Doak, et al. Quantitative conservation biology. *Sinauer, Sunderland, Massachusetts, USA*, 2002.
[39] O. Ovaskainen and B. Meerson. Stochastic models of population extinction. *Trends in ecology & evolution*, 25(11):643–652, 2010.
[40] S. L. Peck. A tutorial for understanding ecological modeling papers for the nonmodeler. *American Entomologist*, 46(1):40–49, 2000.
[41] C. Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.
[42] J. Rousseau. Statistique bayésienne, notes de cours. 2009.
[43] S. Servanty, J.-M. Gaillard, F. Ronchi, S. Focardi, E. Baubet, and O. Gimenez. Influence of harvesting pressure on demographic tactics: implications for wildlife management. *Journal of Applied Ecology*, 48(4):835–843.
[44] N. Sirakov. *Modélisation de la dynamique de population d'une plante native (palmier babaçu) dans le cadre d'un projet de gestion durable au Brésil*. PhD thesis, Université de Montpellier, 2018.
[45] M. Soulé. *Viable Populations for Conservation*. Cambridge University Press, 1987.
[46] S. Tuljapurkar and H. Caswell. *Structured-Population Models in Marine, Terrestrial, and Freshwater Systems*. Population and Community Biology Series. Springer US, 1997.
[47] H. W. Watson and F. Galton. On the probability of the extinction of families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4:138–144, 1875.
[48] T. Wiegand, J. Naves, T. Stephan, and A. Fernandez. Assessing the risk of extinction for the brown bear (ursus arctos) in the cordillera cantabrica, spain. *Ecological monographs*, 68(4):539–570, 1998.