

BERTRAND CLOEZ, TANGUY DAUFRESNE, MARION KERIOUI & BÉNÉDICTE FONTEZ A bayesian version of Galton–Watson for population growth and its use in the management of small population

Volume 12 (2023), p. 49-64. https://doi.org/10.5802/msia.31

© Les auteurs, 2023.

Cet article est mis à disposition selon les termes de la licence CREATIVE COMMONS ATTRIBUTION 4.0. http://creativecommons.org/licenses/by/4.0/



MathematicS In Action est membre du Centre Mersenne pour l'édition scientifique ouverte http://www.centre-mersenne.org/ e-ISSN : 2102-5754

A bayesian version of Galton–Watson for population growth and its use in the management of small population

Bertrand Cloez * Tanguy Daufresne ** Marion Kerioui *** Bénédicte Fontez [‡]

* MISTEA, Université Montpellier, INRAE, Institut Agro, Montpellier, France $E\text{-mail}\ address:\ bertrand.cloez@inrae.fr$

** Eco&Sol, Université Montpellier, INRAE, Institut Agro, Montpellier, France $E\text{-mail}\ address:$ tanguy.daufresne@inrae.fr

*** MISTEA, Université Montpellier, INRAE, Institut Agro, Montpellier, France [‡] MISTEA, Université Montpellier, INRAE, Institut Agro, Montpellier, France *E-mail address*: benedicte.fontez@supagro.fr.

Abstract

Sharp prediction of extinction times is needed in biodiversity monitoring and conservation management. 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. Parameters of this model can be estimated through the Bayesian inference framework. This enables to consider various scenarios. 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) An approach which leads to more information with less computing iv) inclusion of expert or previous knowledge...It can be seen to go one step further than the classical matrix population model for the viability problem. To illustrate these features, we provide analysis details for a real life example with French Pyrenean brown bears.

1. Introduction

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?

The simulations are frequently used to work on these questions even if some works highlight the use of Mathematical results for these problems as [11, 15, 20, 25, 32] among many others.

One of the most famous and powerful mathematical settings for Population Viability Analysis (PVA) is the Matrix population model. They were introduced by [3, 26, 27] and they have become a classical tool to predict the evolving size of a population as illustrated in the books [11, 39]. When the environment is assumed unchanged/stable, some issues listed hereafter

B.C. has been supported by ANR projects, funded by the French Ministry of Research ANR NOLO (ANR-20-CE40-0015) and from the Chair "Modélisation Mathématique et Biodiversité" of VEOLIA-Ecole Polytechnique-MnHn-FX..

Keywords: Bayesian, Galton-Watson, Population Growth, Small population.

are not currently being resolved successfully. We reformulated those issues in terms of elementary questions.

- (I) What are the chances on saving a small population? How probable is extinction? For how long can the small population survive?
- (II) What is the expected short-time evolution of the numbers of individuals?
- (III) How can we proceed to avoid extinction [12, 13], to maintain a population size [36] or wipe out some population [5, 17, 28]. For instance, how many re-introductions are needed to reach a threshold of viability for the population with high probability?

We propose a reformulation of the matrix population model to be viewed as a Bayesian version of a Galton–Watson modelling that allows to address these issues.

2. Bayesian Modelling of Growth Population

2.1. Matrix Population Model

The state of the population, at a certain time $t \in \mathbb{N} = \{0, 1, ...\}$, is given by a (column) vector, $N(t) = (N_1(t), ..., N_K(t))^{\intercal}$, 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,

$$N_j(t+1) = \sum_{i=1}^{K} A_{i,j} N_i(t).$$
(2.1)

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 stage or location), and A then codes the development (or the migration), the birth and the death. The matrix A is sometimes partitioned into two sub-matrices of probabilities: one for transitionsurvival (*TS*) plus one for recruitment (R) : $N_j(t+1) = \sum_{i=1}^{K} TS_{i,j}N_i(t) + \sum_{i=1}^{K} R_{i,j}N_i(t)$. In this work, A includes the whole (transition, survival and recruitment).

2.2. Multiple Galton–Watson as a Matrix Population Model

Multi-type Galton–Watson processes, were introduced in [4, 40], and can be seen as a matrix population model with random number of offspring (but with non-random environment). In this approach, at each time t, each individual l of type i gives a discrete random number $\xi_{i,j,l,t} \in \{0, 1, 2, \ldots, \kappa_{i,j}\}$ of children of type j. Here $\kappa_{i,j} \in \mathbb{N} \setminus \{0\}$ represents the maximal number of offspring (of type j from an individual of type i); it is assumed to be finite. It is also assumed a stationarity of the process (i.e. parameters are not time varying) and independence between observation times and individuals. Denote $N_{i,j}(k,t) = \sum_{l=1}^{N_i(t)} \mathbb{1}_{\{\xi_{i,j,l,t}=k\}}$, then for $t \in \{0, \ldots, T\}$, these assumptions imply :

$$(N_{i,j}(0,t+1),\ldots,N_{i,j}(\kappa_{i,j},t+1)) \sim \text{Multinomial}(N_i(t),p_{i,j}),$$
(2.2)

where probabilities $p_{i,j} = (p_{i,j}(0), \ldots, p_{i,j}(\kappa_{i,j}))$ represent the distribution law for the progeny of type *j* from an individual of type *i*. More precisely, an individual of type *i* will give, after one unit of time, *k* individuals of type *j* (including potentially himself), with probability $p_{i,j}(k)$. We further assume that all sampling in (2.2) are independent. For simplicity, formula (2.1) turned into

$$N_j(t+1) = \sum_{i=1}^K \sum_{k=1}^{\kappa_{i,j}} k N_{i,j}(k,t).$$
(2.3)

Of course setting

$$A_{i,j}(t) = \frac{1}{N_i(t)} \sum_{k=1}^{\kappa_{i,j}} k N_{i,j}(k,t),$$

we recover the model matrix form of equation (2.1) 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 [39, 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 [15] and references therein.

2.3. Prior of the Multiple Galton–Watson model

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. This prior will be the convolution of Dirichlet distribution [34, Section A.8 p. 521]; namely $p_{i,j}$ are independent from each other and we have:

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

$$(2.4)$$

This is a classical law on discrete probability. Parameters $(\alpha_{i,j})$ are called hyper parameters and correspond to the information to incorporate in the estimation. They are fixed by the users and can take into account previous studies or expert knowledge. Among all the choices, three main strategies can be chosen.

- (1) 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 [34, Section 3.2.3] and [35, Section 6]). Roughly, it is the more random choice or the less informative one.
- (2) 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 [12, 13].

If 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

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

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

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

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

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.

(3) Finally, the expert choice can be driven by external data and the weight of the expertise could be automatic. Indeed, Parameters $(\alpha_{i,j})$ can incorporate belief which does not come from any data. They can represent the balance between data and expertise. If the experts think that

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

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

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

2.4. Posterior

We assume that our demographic data has the form of life tables $(n_{i,j}(k,t))_{1 \le i,j \le K,k,t \ge 0}$; 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. We assume to know all the population information over the time interval $\{0, \ldots, T\}$; namely our data is $\mathbf{n} = (n_{i,j}(k,t))_{1 \le i,j \le K,k \ge 0,T \ge t \ge 0}$.

Due to the well known conjugation property between multinomial and Dirichlet distribution (see for instance [34, 35]), and the form of the Galton–Watson transitions and the Markov property, the form of the posterior is simple (see [30, Section 3] and [19, p. 1284] for a detailed computation). More precisely, for every i, j, the posterior of the parameter $p_{i,j}$ conditioned on the data **n** are

$$\mathcal{L}(p_{i,j} | \mathbf{n}) = \text{Dir}\left(\alpha_{i,j}(0) + \sum_{t=0}^{T} n_{i,j}(0,t), \dots, \alpha_{i,j}(\kappa_{i,j}) + \sum_{t=0}^{T} n_{i,j}(\kappa_{i,j},t)\right).$$
 (2.5)

This posterior sums up all the information we have coming from the data and our population model. To answer our main questions, it remains to summarize this object to the principal objects of interest (chance of extinction, extinction time, abundance, etc.). According to the objectives of the outcomes, many choices of modelling are over possible and we will describe three different ones: use of posterior estimators for $p_{i,j}$ (mean or mode), use of posterior simulations of $p_{i,j}$ to construct a scenario approach and use of the posterior to compute analytical integration of $p_{i,j}$.

3. Viability and Small Population Management

Population viability analysis (PVA) [6, 7, 31, 33, 38] aims to manage or predict population development from some ecological, genetic or demographic assumptions. It generally relies on modelling purposes. Assuming a Bayesian Galton–Watson model, posterior estimation or prediction can be used to construct a pathway to PVA.

3.1. Random Mean Matrix as a generalisation of the Population Projection Matrix

An important quantity associated with Galton–Watson model is the random mean matrix M whose entries are

$$M_{i,j} = \sum_{k=0}^{\kappa_{i,j}} k p_{i,j}(k).$$
(3.1)

It is a random matrix since $p_{i,j}$ are supposed to be random in the Bayesian framework. This matrix is the counterpart to the population projection matrix in matrix population model [11, 39].

In the particular case where types correspond to ages or stages (and individuals can only get old or reproduce), then we recover that M is the classical Leslie matrix.

Under general assumptions [21, 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

$$uM = \lambda M, \ Mv = \lambda v \tag{3.2}$$

and

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

As explained in detail in [34, Section 4], the posterior mean, the posterior mode, the posterior median, *etc.* can be used to build an estimator of M. Note that some of these types of estimator can be easily computed; for instance

$$\mathbb{E}[M_{i,j} \mid \mathbf{n}] = \frac{\sum_{k=1}^{\kappa_{i,j}} k\left(\alpha_{i,j}(0) + \sum_{t=0}^{T} n_{i,j}(k,t)\right)}{\sum_{k=0}^{\kappa_{i,j}} \left(\alpha_{i,j}(0) + \sum_{t=0}^{T} n_{i,j}(k,t)\right)}.$$
(3.3)

The main drawback of this approach is that it loses the error estimation which is naturally embedded in the posterior. However, using punctual estimation allows to compute λ and gives an estimation of the viability. Another way is to compute the *Probability of viability* from posterior simulation:

Do $n_{\rm prec}$ 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 (3.1).

Then count the proportion of λ being larger than 1.

3.2. Probability of Extinction

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

$$\mathbb{P}(\text{Ext} \mid N_1(0), \dots, N_K(0), (p_{i,j})) = s_1^{N_1(0)} \times \dots \times s_K^{N_K(0)},$$
(3.4)

where $s = (s_1, \ldots, s_K)$ is the unique solution of $s = \varphi(s)$ in the simplex $\Delta = \{s \in [0, 1]^K \mid \sum_{i=1}^K s_i = 1\}$. Function $\varphi = (\varphi_i)_{1 \le i \le K}$ is the generating function associated with p; for $i \in \{1, \ldots, K\}$, it is defined by

$$\varphi_i(s_1, \dots, s_K) = \sum_{i_1, \dots, i_k} s_1^{i_1} p_{i,1}^{i_1} \dots s_K^{i_K} p_{i,K}^{i_K}.$$
(3.5)

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

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

The variable s_i is the probability of extinction starting from one individual with type *i*.

The posterior distribution of $(p_{i,j})$ can be used to simulate the distribution of the probability of extinction or to compute punctual estimation of this probability of extinction at different quantiles of the posterior.

In practice, to compute the probability of extinction from posterior simulation:

Do $n_{\rm prec}$ 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 (3.5). 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.

Another option is to calculate $\mathbb{P}(\text{Ext} | (n))$, directly integrating Equation (3.4) under the Dirichlet distribution given in Equation (2.5). 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.

3.3. Time to Extinction

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 $t \ge T_{\text{Ext}}$ (and $N(T_{\text{Ext}} - 1) > 0$). Even if one can estimate T_{Ext} by simulation, sharp analytic bounds have been proved. For instance,

$$\mathbb{P}\left(T_{\text{Ext}} > t \mid N_{1}(0), \dots, N_{K}(0), (p_{i,j})\right) = \mathbb{P}\left(\sum_{i=1}^{K} N_{i}(t) > 0 \mid N_{1}(0), \dots, N_{K}(0), (p_{i,j})\right) \\
\leq \lambda^{t} \frac{\sum_{i=1}^{K} v_{i} N_{i}(0)}{\min_{1 \le i \le K} v_{i}}.$$
(3.6)

See for instance [21, Equation (5.54)]. This bound allows to find an upper bound of the extinction time with high probability. To our knowledge, there was no equivalent lower bound in great generality. However, [21, 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 6.1, we generalize this result for K > 1 and show how to use it and (3.6) to bound the extinction time.

In practice, to bound the extinction time T_{Ext} between two values

$$T_{-} \leq T_{\rm Ext} \leq T_{+}$$

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

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

where

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

where v is given in Equation (3.2).

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 (3.6) and Lemma 6.1. 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 $\lambda < 1$.

3.4. Further results

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 (3.5). To do this step one can use any classical optimization algorithm.

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

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 (3.4) and Formula (3.6) hold when replacing N(0) by N(T) and conditioning on N(T) > 0. If we know that some individuals remain at time T, then 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 its associated deterministic model up to error around $\sqrt{N(t)}$. Consequently for short time projection, matrix population model and Galton–Watson should give similar results.

All of these mathematical results can be found in [1, 21, 22, 24, 29].

4. Real data: French Pyrenean brown bears

4.1. Context

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. A previous study [13] established that "successful conservation strategies for this population would require releasing more bears in both sub-populations in the near future". The authors used a deterministic matrix model for estimation and stochastic simulation for prediction. Let us give here a short study on the powerful properties of our Bayesian approach which used a stochastic matrix model and posterior distribution for both estimation, prediction and simulation.

The data we used comes from [9]. It corresponds to the exhaustive supervision of all the population between 2005 and 2016 of the French Pyrenean brown bears. This population is split

Parameter	Posterior	Mean	IC 90%
$p_{1,2}(1)$	Beta (18,4)	0.8181	[0.67, 0.93]
$p_{2,3}(1)$	Beta(17,1)	0.9444	[0.84,1]
$p_{3,4}(1)$	Beta(13,3)	0.8125	[0.64, 0.94]
$p_{4,5}(1)$	Beta(13,1)	0.9286	[0.79,1]
$p_{5,5}(1)$	Beta(73,4)	0.9480	[0.90, 0.98]
$p_{5,1}$	Dir(71, 9, 7, 1)		
$p_{5,1}(0)$	Beta(71,17)	0.8068	[0.73, 0.87]
$p_{5,1}(1)$	Beta(9,79)	0.1023	[0.05, 0.16]
$p_{5,1}(2)$	Beta(7,81)	0.0795	[0.04, 0.13]
$p_{5,1}(3)$	Beta(1,87)	0.0114	[0,0.03]

TABLE 4.1. Posterior for all parameters $p_{i,j}(k)$ and the vector valued $p_{5,1}$ for the bears population example. We use the data [9] and the non-informative prior.

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.

4.2. Modelling assumptions

We consider the same structure model as [13]. 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 $i \in \{1, \ldots, 4\}$, 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 [13] for biological motivations.

We assume that surviving and reproducing are two independent events even if generally it is assumed that reproduction is only possible for surviving parent. Both choices are modelling 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 parameters 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.

4.3. **Prior**

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 $k \in \{0, ..., 3\}$, $\alpha_{5,1}(k) = 1$.

4.4. Results

Using Formula (2.5), the posterior law is easily calculable from the data of [9] and is then given in Table 4.1. The Beta distribution can be seen as a particular case of Dirichlet distribution.

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 knowledge of the previous years $2005, \ldots, t-1$ (that is we do not use all the data set to estimate $(p_{i,j})$). We represent these successive predictions in Figure 4.1. Each boxplot represents the distribution of the prediction.

Small Population Management



FIGURE 4.1. 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 (3.3).

The mean estimator given by Formula (3.3) 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 posterior estimation of $(p_{i,j})$ one can find

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

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 (3.4) have no explicit solutions but are functions of $p_{i,j}$ and one can use simulated values from the posterior 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 [9].



FIGURE 4.2. Density of parameter s_1, \ldots, s_5 for the bear example. Data comes from [9] and we use a non-informative prior.

However one can still see in Figure 4.2 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 (3.4), 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.

4.5. Sensitivity to the Prior

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 [13]. The data they used come from different articles, where the full description of the estimation procedure is not always available. If we look at the infantile mortality rate $p_{1,2}(0)$ then it was based on [41]. 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 Beta(18.2417, 3.7855). This is very close to the non-informative estimation Beta(18, 4) given in Table 4.1. Hence, even if the sample size data seems limited, there is enough information to limit the sensitivity of the prior. We then restrict ourselves to the non-informative prior because it gives similar results and several incomes are unknown as for instance for choosing $(\alpha_{1,5})$.

4.6. Missing Data or Information

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 $n_{5,1}(\cdot, t)$ for the last year of observation, and these data were not used in the previous estimation. However, one can slightly change the Bayesian model to take into account this missing information. Indeed, one can add some structure and assume that

$$p_{5,1}(k) = \sum_{l \ge k} q(l) \binom{l}{k} p^k (1-p)^{l-k}$$

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 [9], 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 [13]. 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.

5. 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 [8, 15, 18, 23, 32]. Deterministic models including differential equation and matrix products are well suited to large population (but sometimes more than 10^6 individuals [10]). 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 small data sets, our approach takes into account demographic stochasticity and enables to add expert knowledge in the prior (via ($\alpha_{i,j}(k)$) setting) and improves posterior estimation and prediction. The Bayesian setting we present can also be adapted for missing data or less informative data. In the use case, in addition to the viability study, we can integrate the sex ratio of the bear cubs as a latent variable in the model and estimate its value as 0.583.

The combination between Galton–Watson processes and Bayesian, and known mathematical results [1, 21, 22, 29, 30] 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 made for small data but easily scalable to big data : in contrast with ABC methods [16] or other MCMC methods [37], 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 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 approach we present offers the advantage of taking into account all the information of an exhaustive dataset to combine estimation, prediction and simulation in one unify approach, conditioned on the data (posterior), and using the expert knowledge (prior). All the results are calculated or simulated from the posterior distribution (a Dirichlet distribution). In the use case of French Pyrenean brown bears, the posterior distribution is given in Table 4.1 and the viability analysis based on this posterior indicates that the reproduction capacity seems to

guarantee viability. Those results come from a Bayesian approach and are therefore conditioned by the data and the model assumptions.

This modelling 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. This model 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. Genetic aspects that can be of importance for small populations is also 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 [14, 22].

An other assumption that we have made for bears example, 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. From [30], we can extend Formula (2.5) for the following setting: for every k_1, \ldots, k_K , we replace all $p_{i,j}(k)$ by $p_i(k_1, \ldots, k_K)$ 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 $\prod_{j=1}^{K} \{1, \ldots, \kappa_{i,j}\}$ instead of $\{1, \ldots, \kappa_{i,j}\}$) 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 of our approach, we can suppose that offspring are 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 [2]). Also the naturally associated prior will be the Gamma law. It also verifies properties (conjugation, non-informative, expertise, *etc.*) which ensure an explicit posterior distribution.

6. Appendix

6.1. A lower bound for the extinction time

Lemma 6.1. We have

$$\mathbb{P}\left(T_{\text{Ext}} \ge t\right) \ge \frac{\min(v)^2}{\max(v)^2} \frac{\lambda^{2t} \left(\sum_{j=1}^K v_j N_j(0)\right)^2}{\Xi \lambda^{t-1} \left(\frac{1-\lambda^t}{1-\lambda}\right) \left(\sum_{j=1}^K v_j N_j(0)\right) + \lambda^{2t} \left(\sum_{j=1}^K v_j N_j(0)\right)^2},\tag{6.1}$$

where

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

Moreover, the right-hand side of equation (6.1) is equivalent, when $t \to \infty$, to

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

Proof. The proof is inspired from [21, Box 5.2 p. 119] We have

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

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 that

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

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

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

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

$$M_{i,j,j'}^{(2)} = \mathbb{E}\left[\left(\zeta_{i,j,l,t}\right)\left(\zeta_{i,j',l',t}\right)\right].$$

If $j \neq j'$ then, under our assumptions, we have

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

and either

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

Then

$$\sum_{j,j'=1}^{K} v_j v_{j'} \left(\sum_{i=1}^{K} N_i(t) M_{i,j,j'}^{(2)} + \sum_{i,i'=1}^{K} N_i(t) N_{i'}(t) M_{i,j} M_{i',j'} - \sum_{i=1}^{K} N_i(t) M_{i,j} M_{i,j'} \right)$$
$$= \lambda^2 X(t)^2 + \sum_{j,j'=1}^{K} v_j v_{j'} \left(\sum_{i=1}^{K} N_i(t) M_{i,j,j'}^{(2)} - \sum_{i=1}^{K} N_i(t) M_{i,j} M_{i,j'} \right)$$
$$= \lambda^2 X(t)^2 + \sum_{j=1}^{K} \sum_{i=1}^{K} N_i(t) v_j^2 \left(M_{i,j,j}^{(2)} - M_{i,j}^2 \right) = \lambda^2 X(t)^2 + \sum_{j=1}^{K} \sum_{i=1}^{K} N_i(t) v_j^2 \sigma_{i,j}^2.$$

Finally, we obtain

$$\mathbb{E}[X(t+1)^2] \le \lambda^2 \mathbb{E}[X(t)^2] + \Xi \mathbb{E}[X(t)]$$

$$\le \lambda^2 \mathbb{E}[X(t)^2] + \Xi \lambda^t X(0)$$

$$\le \Xi \lambda^t \left(\frac{1-\lambda^{t+1}}{1-\lambda}\right) X(0) + \lambda^{2(t+1)} X(0)^2$$

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

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

Acknowledgements

The authors are grateful to the anonymous referees for their comments.

References

- [1] Krishna B. Athreya and Peter E. Ney. Branching processes. Courier Corporation, 2004.
- [2] Andrew D. Barbour, Lars Holst, and Svante Janson. Poisson Approximation. Oxford Science Publications. Clarendon Press, 1992.
- [3] Harro Bernardelli. Population waves. J. Burma Res. Soc., 31:3–18, 1941.
- [4] Irénée-Jules Bienaymé. De la loi de multiplication et de la durée des familles. Soc. Philomat. Paris Extraits, Sér, 5:37–39, 1845.
- [5] 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, 2015.
- [6] Mark S. Boyce. Restitution of gamma-and K-Selection as a Model of Density-Dependent Natural Selection. Annu. Rev. Ecol. System., 15(1):427–447, 1984.
- [7] Mark S. Boyce. Population Viability Analysis. Annu. Rev. Ecol. System., 23(1):481–497, 1992.
- [8] B. J. Cairns, J. V. Ross, and T. Taimre. A comparison of models for predicting population persistence. Ecol. Model., 201(1):19–26, 2007. Management, Control and Decision Making for Ecological Systems.
- [9] J. J. Camarra, J. Sentilles, A. Gastineau, and P. Y. 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.
- [10] F. Campillo and C. Lobry. Effect of population size in a predator-prey model. Ecol. Model., 246:1–10, 2012.
- [11] Hal Caswell. Matrix population models: construction, analysis, and interpretation. Sinauer Associates, 1989.
- [12] Hal Caswell, Masami Fujiwara, and Solange Brault. Declining survival probability threatens the North Atlantic right whale. Proc. Natl. Acad. Sci. USA, 96(6):3308–3313, 1999.
- [13] Guillaume Chapron, Pierre-Yves Quenette, Stéphane Legendre, and Jean Clobert. Which future for the French Pyrenean brown bear (Ursus arctos) population? An approach using stage-structured deterministic and stochastic models. C. R. Biol., 326:174–182, 2003.
- [14] Daryl J. Daley. Extinction conditions for certain bisexual Galton-Watson branching processes. Z. Wahrscheinlichkeitstheor. Verw. Geb., 9(4):315–322, 1968.
- [15] Steinar Engen, Russell Lande, Bernt-Erik Sæther, and Henri Weimerskirch. Extinction in relation to demographic and environmental stochasticity in age-structured models. *Math. Biosci.*, 195(2):210– 227, 2005.

- [16] Arnaud Estoup, Eric Lombaert, Jean-Michel Marin, Thomas Guillemaud, Pierre Pudlo, Christian Robert, and Jean-Marie Cornuet. Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Mol. Ecol. Resour.*, 12(5):846–855, 2012.
- [17] Esther Gerber, Christine Krebs, Craig Murrell, Marco Moretti, Remy Rocklin, and Urs Schaffner. Exotic invasive knotweeds (Fallopia spp.) negatively affect native plant and invertebrate assemblages in European riparian habitats. *Biol. Conserv.*, 141(3):646–654, 2008.
- [18] Lev R. Ginzburg, Lawrence B. Slobodkin, Keith Johnson, and Andrew G. Bindman. Quasiextinction Probabilities as a Measure of Impact on Population Growth. *Risk Analysis*, 2(3):171–181, 1982.
- [19] Miguel González, Jacinto Martín, Rodrigo Martínez, and Manuel Mota. Non-parametric Bayesian estimation for multitype branching processes through simulation-based methods. *Comput. Stat. Data Anal.*, 52(3):1281–1291, 2008.
- [20] Frédéric Gosselin and Jean-Dominique Lebreton. Potential of branching processes as a modeling tool for conservation biology. In *Quantitative methods for conservation biology*, pages 199–225. Springer, 2000.
- [21] Patsy Haccou, Patricia Haccou, Peter Jagers, and Vladimir A. Vatutin. Branching processes: variation, growth, and extinction of populations. Cambridge University Press, 2005.
- [22] Theodore E. Harris. The theory of branching processes. Courier Corporation, 2002.
- [23] Elizabeth Eli Holmes, John L. Sabo, Steven Vincent Viscido, and William Fredric Fagan. A statistical approach to quasi-extinction forecasting. *Ecol. Lett.*, 10(12):1182–1198, 2007.
- [24] Marek Kimmel and David E Axelrod. Branching Processes in Biology, volume 19 of Interdisciplinar Applied Mathematics. Springer, 2002.
- [25] Jean-Dominique Lebreton, Frédéric Gosselin, and Colin Niel. Extinction and viability of populations: paradigms and concepts of extinction models. *Ecoscience*, 14(4):472–481, 2007.
- [26] Patrick H Leslie. On the use of matrices in certain population mathematics. *Biometrika*, pages 183–212, 1945.
- [27] E. G. Lewis. On the Generation and Growth of a Population. Sankhyā, Ser. A, 6(1):93–96, 1942.
- [28] Sarah Lowe, Michael Browne, Souyad Boudjelas, and Maj 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.
- [29] Sylvie Méléard. Modèles aléatoires en Ecologie et Evolution. Springer, 2016.
- [30] Manuel Mendoza and Eduardo Gutiérrez-Peña. Bayesian conjugate analysis of the Galton–Watson process. Test, 9(1):149–171, 2000.
- [31] William F. Morris, Daniel F. Doak, et al. Quantitative conservation biology. Theory and Practice of Population Viability Analysis. Sinauer, 2002.
- [32] Otso Ovaskainen and Baruch Meerson. Stochastic models of population extinction. Trends Ecol. Evol., 25(11):643–652, 2010.
- [33] Steven L. Peck. A Tutorial for Understanding Ecological Modeling Papers for the Nonmodeler. Am. Entomol., 46(1):40–49, 2000.
- [34] Christian Robert. The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer, 2007.
- [35] Judith Rousseau. Statistique Bayésienne, Notes de cours. http://www.crest.fr/ckfinder/ userfiles/files/pageperso/mandre/sb_cours.pdf, 2009.
- [36] Sabrina Servanty, Jean-Michel Gaillard, Francesca Ronchi, Stefano Focardi, Eric Baubet, and Olivier Gimenez. Influence of harvesting pressure on demographic tactics: implications for wildlife management. J. Appl. Ecol., 48(4):835–843, 2011.

- [37] Nikolay 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.
- [38] M. Soulé. Viable Populations for Conservation. Cambridge University Press, 1987.
- [39] Shripad Tuljapurkar and Hal Caswell. Structured-Population Models in Marine, Terrestrial, and Freshwater Systems. Population and Community Biology Series. Springer, 1997.
- [40] Henry William Watson and Francis Galton. On the probability of the extinction of families. *Journal* of the Anthropological Institute of Great Britain and Ireland, 4:138–144, 1875.
- [41] Thorsten Wiegand, Javier Naves, Thomas Stephan, and Alberto Fernandez. Assessing the risk of extinction for the brown bear (Ursus arctos) in the Cordillera Cantabrica, Spain. Ecol. Monogr., 68(4):539–570, 1998.