

Bayesian Models for Pest Detection

Chris Malone

2022-03-28

Contents

Draft Outline	5
1 Introduction	7
An unnumbered section	7
2 Literature review	9
2.1 Inferring absence of cryptic pests	9
2.2 Standard approaches to inferring absence of a cryptic species . .	9
2.3 [Introduce most recent work] Incorporating population growth and Bayesian posterior inference	11
2.4 [Motivate my extension] Limitations of the basic Bayesian model	12
3 A generalised model of pest species detection	15
3.1 Introduction	15
3.2 Agent based simulation (ABS)	15
3.3 Approximate Bayesian computation (ABC)	15
3.4 Priors	15
3.5 Model specification	16
3.6 Sampling	16
3.7 A simple example	16
4 Case study: <i>Ceratatis capitata</i>	17
4.1 Example case: Medfly	17
4.2 [Motivating the model extension] Problems posed by Medfly for the Barnes model	17
4.3 Full model specification	18
4.4 Citations	18

Draft Outline

The thesis will be composed of the following sections:

- Abstract (120-200 words; write last)
 - Motivation
 - Research problem
 - Proposed solution
 - Application
- Table of Contents
 - Main sections should be numbered, but references and appendices should not be.
- Introduction
 - Establish broad background (real world motivation)
 - Establish need for further work
 - Discuss the contribution of the thesis
- Literature review
 - Covering theoretical background, including
 - * Nature of the pest detection problem
 - * Attempts to solve the problem
 - Frequentist method
 - Barnes et al.'s model
- Extended model
 - Discussion of the extended model, broken into major components, i.e.
 - * Probabilistic graphical model
 - * Priors
 - * Sampling
- Case study
 - Background
 - Model setup
 - Results

Chapter 1

Introduction

An unnumbered section

Chapter 2

Literature review

In this chapter, I review some literature relating to the core problem that I am concerned with in this thesis. I begin by discussing the general problem of inferring absence of a cryptic, incipient pest species population. In particular, I emphasise that the problem requires the marriage of separate mathematical modelling cultures; (a) mathematical/ecological modelling, and (b) statistical inference.

2.1 Inferring absence of cryptic pests

The primary problem this work is concerned with is to infer the absence of a cryptic pest species. A cryptic species is any species which is hard to detect. One example is snakes, which are small, immobile, camouflaged, and located in inaccessible habitats (Inferring the Absence of a Species: A Case Study of Snakes, Kery 2002). A problem that environmental managers and governors face is to determine the minimal length of time required for. I will refer to this problem as the **problem of post-outbreak surveillance program design**, or, alternatively, as the problem of program design.

2.2 Standard approaches to inferring absence of a cryptic species

In this section, I outline the standard existing approach to inferring pest absence. The simplest approach to this problem is given by McArdle (1990). Let the *rarity* of the species $p \in [0, 1]$ be the probability that a species is detected in any given sampling unit. (Sampling units can be arranged spatially or temporally; e.g. a survey that involves checking w weeks at k locations would have wk sampling units.) Then, the number of surveys in which the species is detected is given by $X \sim \text{Binomial}(T, p)$. Accordingly, the probability of *not* detecting the species

in T surveys is given by

$$\alpha = \Pr(X = 0) = 1 - (1 - p)^T.$$

The last formula allows us to compute any of the 3 quantities α , p , and t , assuming the other two are given.

Given this framework, the problem of planning program becomes the following. We decide *a priori* what the smallest “rarity” p worth detecting is. McArdle supposes that if a species is sufficiently difficult to detect (while, nonetheless, $p > 0$) then it cannot be considered a member of an ecological community, and therefore not worthy of being deemed “present”. Write the smallest rarity worth detecting as p_0 . Then, we choose a minimum detection probability α that we are willing to accept. For example, we might wish to have a chance of $\alpha > 0.95$ of detecting a species, given that it is present. Then, with the above formula, we can rearrange to get the smallest number of survey units T such that detection probability α is achieved.

Statisticians will recognise that the above is essentially power analysis for data modelled as identically and independently distributed Bernoulli trials. This analogy can be made more concrete. For any fixed rarity p , we can derive the probability of observing n or more negative surveys. This is the p-value. We can then reject the hypothesis that there is the rarity is greater than p_0 , the rarity worth detecting.

McArdle’s method is popular in the literature, most likely due to its conceptual simplicity.

1996. Using statistical probability to increase confidence of inferring species

2.2.1 Limitations of the standard approach

So far, I have discussed a general model, first proposed by McArdle (1990), for surveillance of a species in general. This model works well for a range of cases. However, for pest surveillance, this model has several limitations. Firstly, for monitoring of cryptic pest populations the minimal permissible rarity p_0 may be arbitrarily small. For example, this may be the case if (a) a pest is extremely hard to detect, so that even moderate sized populations have low detection probability, and/or (b) pests have high invasive potential, so that even extremely small populations pose an invasive threat. However, when $p_0 = 0$, the probability of non-detection is 1 at each timestep.

The model extracts information from the

1. No probabilities
 - Probabilities help decision makers
2. No estimation for p .

2.3. [INTRODUCE MOST RECENT WORK] INCORPORATING POPULATION GROWTH AND BAYESIAN PO

- For cryptic pest species, the model is even for dangerous levels of p .
3. No model of growth.
 4. Not suitable for arbitrarily small p_0 .
 - There may be no $p_0 > 0$ that we would consider acceptable. Then, the goal is to reject $\mathbb{H}_0 : p \leq p_0 = 0$. The problem is that the model is that non-detections have probability 1 under the null hypothesis that $p_0 \leq 0$. I.e. the probability of detecting a member of the species is at most 0 under this setting. In other words, $\alpha_n = 0$, for any number of surveys n . To get around this, we might choose p_0 to be arbitrarily small. However, such a choice would be arbitrary and not motivated by scientific theory or value-judgment considerations. Also, setting p_0 to be arbitrarily small may force us to be overly conservative. The smaller that we set p_0 , the larger the number of surveys n required to reject $\mathbb{H}_0 : p \geq p_0$.
 5. No prior information.
 - This is related to the previous point. The above does not allow us to incorporate prior information about the rarity of the species. If we have reason to believe that the species is nearly extinct, then the above methodology may give overly conservative results. I.e. it might recommend surveying for longer than is required.

2.3 [Introduce most recent work] Incorporating population growth and Bayesian posterior inference

A Bayesian approach to the planning program problem is proposed by Barnes et al. (2021). Barnes extend the model in two basic ways. Firstly, Barnes et al. supplement the detection model with a model of the population dynamics for the species.

2.3.1 Model specification

As mentioned above, the Barnes model consists of (a) a growth model and (b) a detection model. The model for population growth gives us priors for the size of the population at each time and/or location where a survey is completed. In particular, a Poisson-offspring model is assumed for the population size. The Poisson-offspring model is a type of branching process, initially developed by (???). A Poisson-offspring process is a process $X_{t=1}^T$ such that $X_0 \sim \text{Poisson}(\lambda)$, $\lambda \in [0, \infty)$, and $X_t | X_{t-1} \sim \text{Poisson}(X_{t-1}\delta)$. The typical situation that fits this process is the case where the number of individuals at time $t = 0$ is Poisson distributed with mean λ , and each individual born at time t has identically and independently Poisson distributed offspring with mean δ .

2.4 [Motivate my extension] Limitations of the basic Bayesian model

There exists a category of cases under which the assumptions of basic model above are not plausible. In this section, I describe this category of cases, and outline why these cases cause problems for the Barnes model. Three assumptions of the above model are as follows:

1. The growth of an incipient population can be described by a Poisson-offspring model.
2. The probability of detecting a randomly drawn member of the population p is known, or can be reasonably estimated from data.
3. The population grows exponentially at a constant rate. I.e. the growth rate of the population is not random and does not change over time.

When these assumptions are not met, the model may not be a realistic representation of the processes being modelled. Firstly, the Poisson-offspring distribution is most natural when all detectable individuals at time t are offspring of the individuals that were detected at time $t - 1$. However, this will not be the case if members of the population continue to be detectable after giving birth.

It is not always possible to estimate a “rarity” parameter p from data. Recall that the rarity of a population is the probability of detecting a randomly drawn member of the population in a single survey. Thus, it depends fundamentally on the design and intensity of the survey employed. However, survey designs may vary significantly between regions, so that we cannot generalise estimates of p . For example, suppose an environmental manager has estimated p , under a standard, general surveillance program. However, they later suspect an outbreak has occurred, and therefore wish to implement an intensified surveillance program. A concrete example of this can be seen in the case of Mediterranean fruit fly, surveillance is performed using a spatial network of surveillance units (i.e. traps) which are inspected at regular intervals. When an outbreak has occurred, or is suspected, the manager may wish to deploy additional, *supplementary* survey units.

It may be objected that it should be possible to derive the an estimate of the rarity p using the probability-given-distance function p_d . For example, we could simulate

In applied contexts, the probability of capture must be estimated from data. In the case of small insect species, such as Mediterranean fruit fly, probability of capture is estimated with the aid of release-recapture experiments. Sterilised flies are released, in an area containing a standardised grid of surveillance units (traps). Then, after some time has passed, flies are counted. The resulting recapture data lends itself to two natural estimands. Firstly, there is the probability of a randomly selected fly being captured at all, given some trapping grid setup (e.g. a square grid with 400m spacing). This is essentially the “rarity” parameter discussed above. Secondly, there is the probability of a randomly

2.4. [MOTIVATE MY EXTENSION] LIMITATIONS OF THE BASIC BAYESIAN MODEL¹³

selected fly being captured in a given trap, given the distance between the trap and the release location of the flies.

As mentioned above the third assumption is that the population will not be listed.

2.4.1 Summarise

In effect, in cases where the three assumptions listed above do not hold, it will not be ...

Chapter 3

A generalised model of pest species detection

3.1 Introduction

In the previous chapter, I reviewed recent work by Barnes et al. on inferring absence of a cryptic pest species. I concluded the chapter by discussing some of the model’s restricting assumptions, and how they affect the reliability of the model. In the present chapter, I propose to develop a method that relaxes the assumptions listed above, and may work for a relatively broad class of cases. The general proposal is to combine agent based simulation (ABS) with approximate Bayesian computation (ABC) to infer probability of absence when the processes governing population dynamics are arbitrarily complex. I first start by outlining the method, before giving a simple example of the method in action. I finish by discussing some theoretical and practical limitations of the method.

3.2 Agent based simulation (ABS)

3.3 Approximate Bayesian computation (ABC)

3.4 Priors

At the minimum,

The priors are determined by a stochastic model of the population dynamics for the species. Such a model may or may not be “agent based”. For the purpose of this work, I take the term “agent based” to mean that the model takes into account differential individual properties of members of the species. (In this work, I consider differential locations). We model the population dynamics

flexibly as a stochastic process. Simulation allows for the incorporation of an arbitrary degree of biological and ecological complexity into the model. Using an agent based model allows us to define a prior on each individual population member

3.5 Model specification

Under the most general description, the

The basic model is composed of (a) a growth model and (b) a surveillance model. The model may be spatial or non-spatial, depending on. When the model is spatial, it is assumed that we have a stochastic or deterministic function that relates the distance between an individual and a surveillance unit, and that surveillance unit detecting that individual. I.e., we know

3.5.1 The growth model

There are no a priori assumptions on the population dynamics for the growth model. For example, we might apply stochastic or deterministic models of logistic, exponential, or linear growth. The only thing that matters is that we get a distribution over the size of the population at any given time point where surveillance will be conducted.

A natural way to set a prior on the population size at each time point t is to set a prior on the population size at the initial time point, and then assume that the population sizes at other time points are given by some (deterministic or stochastic) function of the population size at $t - 1$, and the value of a covariate vector X_t , which includes variables relevant to population growth.

$$N_t = f(N_{t-1}, X_t)$$

3.5.2 The surveillance (detection) model

It is assumed that surveillances occur at regular time intervals $t \in \{1, \dots, T\}$.

3.6 Sampling

3.7 A simple example

Chapter 4

Case study: *Ceratatis capitata*

4.0.1 Intro: Version 1

In the previous section, a general model for inferring absence/presence of a cryptic, incipient pest species population was introduced. In this chapter, I apply the model to a specific case study.

4.1 Example case: Medfly

An example case of the kind described above can be seen in Mediterranean fruit fly (*C. Capitata*).

4.1.1 Economic importance of the fruit fly

4.1.2 Medfly detection problems

Medfly have low dispersal. This means that flies may go undetected across generations. <https://onlinelibrary-wiley-com.virtual.anu.edu.au/doi/pdfdirect/10.1111/j.1570-7458.2006.00415.x>

4.2 [Motivating the model extension] Problems posed by Medfly for the Barnes model

The reader may recall that the model introduced in chapter 3 was motivated by consideration of three restrictive assumptions, listed in chapter 2. These assumptions were...

I will briefly describe why these assumptions do not satisfy.

- Problem 1: Flies live across generations.
- Problem 2: It is hard to estimate a distribution for p .
 - Source for supplementary trapping: https://nre.tas.gov.au/Documents/Review_of_IR_for_FruitFly.pdf
- Problem 3: Growth rate is assumed constant.

The benefit of the model

4.3 Full model specification

As the reader will recall from chapter 3,

4.4 Citations

[Xie, 2015] in the .bib file.

Bibliography

Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. URL <http://yihui.org/knitr/>. ISBN 978-1498716963.