Zero tolerance ecology and the mapping of diseases.

Andrew J. Tyre, Jonathan Rhodes, Tara Martin and Elizabeth VanWormer

Abstract

A fundamental question in ecology broadly has to do with how species are distributed in space, and which biotic and abiotic characteristics of the environment explain that distribution. Veterinary epidemiology could use the same tools to understand environmental drivers of parasites and diseases. Unfortunately, the ecological and epidemilogical data used to build species distribution models suffers from imperfect detection, creating both false presences and false absences. The ecological community developed a wide range of two-stage models to compensate for these errors. Here we explore how those tools could be used by veterinary epidemiologists to map the prevalence of diseases.

Errors arise from a variety of mechanisms in the data generating process. The sources of errors also vary by the scale of inference. We review the recent ecological literature on error processes. Then we examine common types of epidemiological data through this lens to identify errors of the same type as ecology, as well as errors unique to epidemiological studies. For example, pooled samples used to detect diseases in vector species generate a particular kind of prevalence data with a unique error structure.

We review the tools developed by ecologists to mitigate the effects of errors, and consider the applicability of these approaches to epidemiological data. For epidemiologists to apply these tools requires changes in the design of surveys to generate the data necessary to support the two stage models. We finish by describing the opportunities for future work created by the intersection of species distribution modelling and disease mapping.

Introduction

Observations of reality are imprecise and fraught with error from many sources. Arguably, the science of statistics was invented to deal with this fact. In many disciplines, the error in observations can be well described by a Gaussian, or normal, distribution. Statistics is well equipped to deal with those cases! In a few disciplines, errors take on different structures that are not normal (pun intended). Sometimes simply changing the distribution we assume the observations are following is sufficient to regain a powerful suite of tools to draw conclusions. But in a few cases, even that is not sufficient.

Imagine trying to determine the prevalence of a disease in a population out in nature. The most basic observation, is this individual infected (or has been infected in the past), is a simple binomial variable. Even that observation has errors – there are few (maybe no?) methods for detecting disease in a tissue sample of an individual that have no false positive or negatives. But even if our method is perfect, going from a infected proportion of a sample of individuals back to the prevalence in the population is fraught with difficulty. We must assume that the individuals captured are a random sample of the population; that is, infected and uninfected individuals are both randomly distributed and equally likely to be captured. If we are taking samples from across a series of habitat patches in a fragmented landscape, we may also find that no individuals were captured at a particular patch at all. Does that mean the species is absent, and if not, what is the prevalence in that patch? The answers to that final question depend a great deal on the methods of sampling and how sampling effort is distributed in space.

A fundamental question in ecology broadly has to do with how species are distributed in space, and which biotic and abiotic characteristics of the environment explain that distribution. The sub-discipline of "species distribution modelling" is the art of answering this question. Historically, the details of how an observation of a species at a particular point in time and space arose were largely ignored (cite some early paper here). Species distribution models were built assuming that the presence and absence of a species at some series of

points was known with certainty (or cite some examples here). Although this assumption was widely known to be false, there did not seem to be any methods both testing this assumption and mitigating its effects.

The primary issue arising from assuming observations are accurate when they are not is bias (Tyre et al. 2003). If a species is sometimes missed when present, a false negative, estimates of the effects of covariates on the probability of presence are biased towards zero. **is there a paper looking at effects of false positives in a similar broad way? Yes, follow up on blog post from Jonathan.**

Understanding the distribution of diseases is complicated by the fact that the habitat of the disease, the host population, is even more fragmented than the habitat of the host species.

A wildlife disease example of habitat mapping is Chronic wasting disease in deer (Evans et al. 2016). These approaches are being used to prioritize areas for management (citations therein).

Another example is the distribution of different strains of *Toxoplasmosis gondii* across the land-sea interface (VanWormer 2014).

Sources of zeros

Summarize Martin et al. (2005), and update with more recent thinking. Provide some veterinary examples?

What are the questions and what kind of data are collected?

Scale of questions?

how do zeros arise?

how to estimate false negatives? repeat observations?

Pooled sampling of mosquitos - minimum prevalence

infographic / toy examples

What drives detectability in epidemilogical samples and are they different

Tools for mitigating extra zeros

The primary tool for reducing the bias associated with false negatives is to shift the design of a survey so that sample points are surveyed more than once. Even without more complex statistics, this approach reduces bias by providing more opportunities to detect the species at a given sample point. This does induce a severe tradeoff: visiting the same points more than once reduces the number of points that can be sampled if the amount of effort is held constant. It is possible to identify an optimal level for this tradeoff that depends on the goals of the study (Field, Tyre, and Possingham 2005, MacKenzie and Royle (2005)).

In the context of mapping a disease, if the sample point is represented by a population of hosts, then sampling > 1 host per point has the same effect as repeat visits. The same tradeoff also exists – the more individuals sampled at one site the fewer different sites can be sampled with a fixed amount of effort.

Dynamic detection error models –

Joint occurrence / co-occurence SDM

Future directions

Designs accomodating false positive/negative rates at the pointy end

References

Evans, Tyler S., Megan S. Kirchgessner, Brian Eyler, Christopher W. Ryan, and W. David Walter. 2016. "Habitat Influences Distribution of Chronic Wasting Disease in White-Tailed Deer." *The Journal of Wildlife Management* 80 (2): 284–91. doi:10.1002/jwmg.1004.

Field, Scott A, Andrew J Tyre, and Hugh P Possingham. 2005. "Optimizing Allocation of Monitoring Effort Under Economic and Observational Constraints." *Journal of Wildlife Management* 69 (2): 473–82.

MacKenzie, Darryl I, and J Andrew Royle. 2005. "Designing Occupancy Studies: General Advice and Allocating Survey Effort." *Journal of Applied Ecology* 42 (6). Wiley Online Library: 1105–14.

Martin, Tara G, Brendan A Wintle, Jonathan R Rhodes, Petra M Kuhnert, Scott A Field, Samantha J Low-Choy, Andrew J Tyre, and Hugh P Possingham. 2005. "Zero Tolerance Ecology: Improving Ecological Inference by Modelling the Source of Zero Observations." *Ecology Letters* 8 (11). Blackwell Science Ltd: 1235–46.

Tyre, Andrew J, Brigitte Tenhumberg, Scott A Field, Darren Niejalke, Kirsten Parris, and Hugh P Possingham. 2003. "Improving Precision and Reducing Bias in Biological Surveys: Estimating False-Negative Error Rates." *Ecological Applications* 13 (6). Ecological Society of America: 1790–1801.

VanWormer, Melissa A. AND Conrad, Elizabeth AND Miller. 2014. "Using Molecular Epidemiology to Track Toxoplasma Gondii from Terrestrial Carnivores to Marine Hosts: Implications for Public Health and Conservation." *PLOS Neglected Tropical Diseases* 8 (5). Public Library of Science: 1–14. doi:10.1371/journal.pntd.0002852.