An Extension of Two-Dimensional Detection Function Distance Sampling Software

Calliste Fagard-Jenkin

January 2017

1 Introduction

1.1 Context and background

1.1.1 Initial overview

This project focuses on line-transect distance sampling; a hybrid method which uses both design and model based inference to produce estimates on the sizes of animal populations. An extension of distance sampling by Borchers and Cox (2017) allows for the use of 2D detection functions, describing the probability of detection of an animal in both forward and perpendicular directions. This methodology produces estimates of abundance which are far less biased when compared to estimates produced by conventional distance sampling, when the animals in the population of interest display responsive movement. The raison d'etre of this project is to further extend the code-base of the Borchers and Cox (2017) methodology's R package (called LT2D) and to provide a number of illustrative analyses on a range of datasets to demonstrate the improved functionality.

1.1.2 What is distance sampling?

Distance sampling is a methodology within ecological statistics which allows us to produce unbiased estimates of animal population sizes even when the detection of any individual in the survey area is uncertain. First, a quick description of its simpler younger brother, strip sampling (in which detection of individuals is certain), will be given in order to illustrate the fundamental principles of greatest importance.

Consider a survey area within which the population of interest is closed (there is no emigration or immigration of individuals). We further assume that births and deaths are considered to occur at a negligible rate. In this way, the survey is considered to be a snapshot of the population at some given moment. In our example, we choose a square survey area, which we subdivide into five strips of regular width. Animals are uniformly distributed throughout the survey area, in both the perpendicular and the horizontal axis. This is illustrated by figure 1.
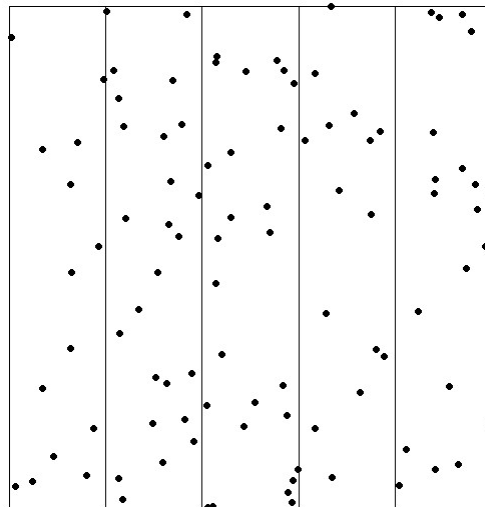

Figure 1: Example distribution of animals in a square survey area divided into 5 equal strips

1.1.3 The significance of 2D detection functions

1.2 Previous contributions

From June - August 2017 I undertook work on the code-base as a summer student. The notable changes I made to this code-base over this period will be clearly outlined in this section to distinguish them from the work which was undertaken from September 2017 onward as part of this project, however these will be not be described in depth, in order to avoid removing any focus from the improvements made as a part of this project. If more detail on the work conducted in this summer studentship is of interest, it can be found in the slides of a presentation conducted at the 2017 St Andrews Mathematics Undergraduate Research Conference, which have been included in the appendix.

1.2.1 Object Oriented Additions

The R programming language allows for packages to include their own classes to produce objects with behaviours and properties tailored to the task at hand. These typically come in the form of S3 or S4 objects. As part of my summer work I wrote methods for S3 generic functions which allow the LT2D package to be more streamlined and provide the user with a cleaner

and more straight forward interface with the functionality provided under the hood. The first set of these methods are dedicated to the R plot function, allowing the user to plot LT2D fitted model objects with a simple call to the generic plot function, with the aforementioned object as the first argument.

1.2.2 Adding in Abundance Estimation

Previous versions of the code-base only produced estimates for average detection probability, $\hat{p}$. Code supplied by Prof. Borchers was modified and integrated with existing LT2D code to allow the package to produce estimates of abundance as well as average detection probability. This work was done towards the end of the project and hence was not fully integrated and tested with all other pieces of functionality available within the software, therefore modifications of abundance estimation code and its continual testing and iterative improvement is necessary for the completion of this project, however the details of these updates are of far less interest than the more ambitious aspects of the project and so they shall not be mentioned further in any significant amount of detail. Any addition or modification of code or theoretical technicalities related to the LT2D package mentioned out-with of section 1.2 are not a result of previous work and form the improvements and extensions which the project aims to produce.

1.2.3 Rounded Angles

When measurements of small detection angles are rounded down to 0 in the field, the magnitude of the bias of estimated average detection probability, $\hat{p}$, can drastically increase. An appropriate modification to the method's likelihood function was theorised by prof. Borchers and then I implemented the new functionality into the source code. I conducted multiple simulations to confirm that bias was significantly reduced, all of which clearly demonstrated the new estimator performed far better.

1.2.4 New Perpendicular Density

A reparameterization of an existing perpendicular density was required to allow for a half-normal perpendicular density function with a non-zero expected quantity of animals on the transect line. Prof. Borchers and I decided on a suitable formulation and I implemented it within the LT2D package.

1.2.5 Concatenation of GoF functions

Goodness of fit tests (Cramer Von Mises and Kolmogorov-Smirnov) are conducted separately in the x and y (perpendicular and forward) directions, respectively, by the LT2D. These were split into distinct functions, making their use tedious to the end user. I rectified this issue by creating a wrapper function which was easier for the user to interact with and acted as an interface with the pre-existing code.

1.2.6 Model Collecting and Summarising

As with many model-based statistical methodologies, it is often good practice to fit a range of varied models to the available data, and to use selection criteria (such as Akaike's Information Criterion) to select which model(s) are most reasonable. This leads to the user's R global environment being peppered with objects created by the LT2D package, making it difficult to keep track of the models of interest. Inspired by another piece of software used for animal abundance (RMARK - An interface to the popular MARK windows software used as a tool for analysis of mark-recapture data), model-collecting and comparison-table-generating functions were implemented within the LT2D package.

1.3 Project objectives and outline

The project aims to include the ability for users to specify covariate inclusion for the two-dimensional detection functions, making them not only a function of perpendicular and forward distance, but also of user-defined relationships between other variables. Upon discussion with Prof. Borchers, it was decided this functionality

should be kept as general as possible, following the following main criteria:

1. Users should be able to choose whether the covariate affects the behaviour of the detection function in the perpendicular direction, the forward direction or at the intercept, or any combination of the above.

2. Users should be able to define the formula which describes the linear predictor for the relevant parameter and the covariate(s) of interest themselves, rather than be restricted to pre-defined options

3. Users should be able to specify different formulas for each parameter (forward, perpendicular and intercept) wherever the detection function allows these to be different.

4. There should be some level of inbuilt error-checking to ensure that users choices for formulas and starting parameter values are consistent between themselves and their choice of detection function

All code calculating estimates of abundance ($\hat{N}$), average detection probability ($\hat{p}$) and AIC should be modified to work with all reasonable choices of covariates and related linear predictors. S3 methods should be written for the plot function to allow it to handle fitted models which contain cases of both discrete and continuous covariate inclusion. The goodness of fit function should also be generalised to use the Kolmogorov-Smirnov and Cramer Von Mises tests when the fitted model has included covariates.

Time permitting, a further objective of the project is to allow the software to cope with a mixture model detection function. This would allow the methodology to cope with populations which are heterogeneous with respect to detection probability. When the survey population consists of distinct sub-populations each with different behaviours which affect detection probability (i.e. males' mating calls make them far easier to detect)