Introduction

Peter Solymos

Point count data analysis workshop, BIOS2 2021, March 16-25

Outline

Day 1

- ~~Introduction~~
- ~~We need to talk about data~~
- ~~A primer in regression techniques~~

Day 2

- ~~Behavioral complexities~~
- ~~Removal models and assumptions~~

Day 3

- The detection process
- Distance sampling

Day 4

- Putting it all together
- Roadside surveys & recordings

Get course materials

1. Visit https://github.com/psolymos/qpad-workshop/releases
2. Download the latest release into a **NEW** folder
3. Extract the zip/tar.gz archive
4. Open the `workshop.Rproj` file in RStudio (or open any other R GUI/console and `setwd()` to the directory where you downloaded the file)
5. Move your **LOCAL** files into the new folder to keep things together

Local copy

Avoid conflicts as we update the workshop materials: work in a **LOCAL** copy of the R markdown files

```
source("src/functions.R")
qpad_local(day=3)
```

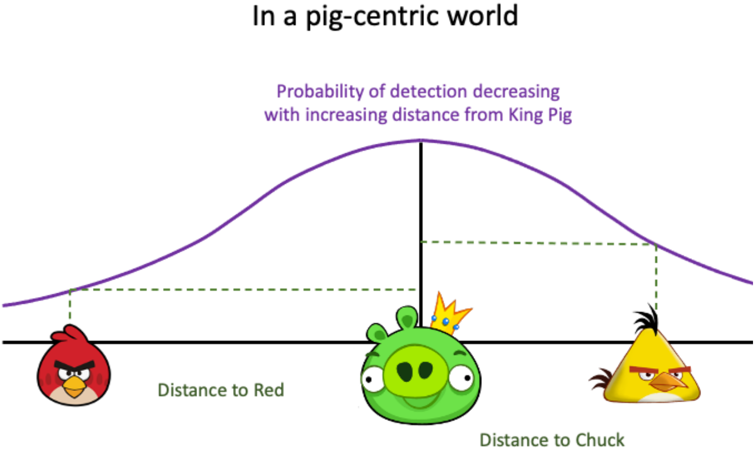
LOCAL copies will not be tracked and overwritten by git. You can copy new files over and it will not impact your local copies.

Estimating nuisance variables

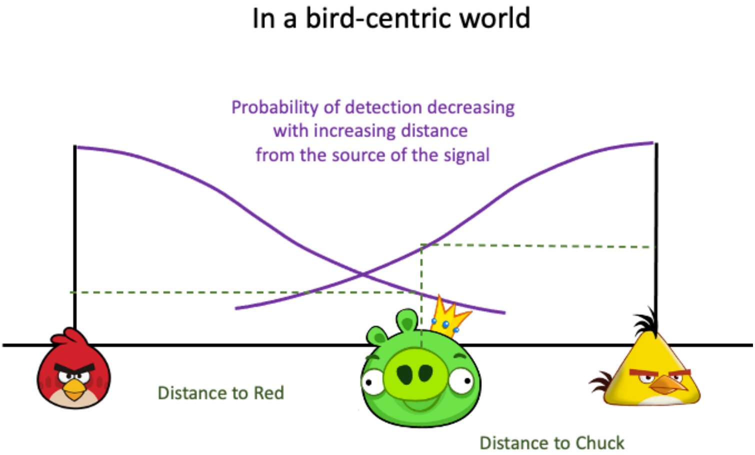
We have discussed how to estimate p based on removal modeling.

Next we will discuss how to estimate q based on distance sampling.

Are we hearing into the forest?



Is the sound coming out of the forest?



The detection process

The detection itself is often triggered by **visual** or **auditory** cues, and thus depend on the individuals being available for detection (and of course being present in the survey area).

The detection is the physical act of registering a signal, and is related to sensitivity of the receptor, and not the psychological processing of the detected signal (detection vs. transcription). I.e. detection depends on the distance, but not perceived distance (measurement error).

The distance function

$g(d)$ describes the probability of detecting an individual given the distance d between the **source** of the signal and **receptor** (e.g. human ear).

- It is a monotonic decreasing function of distance,
- $g(0) = 1$: detection at 0 distance is perfect.

Negative Exponential

- one-parameter function: $g(d) = e^{-d/\tau}$
- probability quickly decreases with distance, this mirrors sound attenuation under spherical spreading
- suitable form for acoustic recording devices,
- not a very useful form for human based counts

Half-Normal

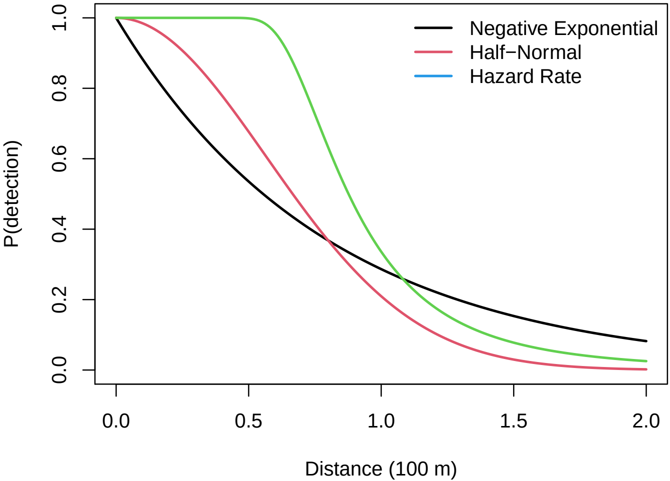
- one-parameter function: $g(d) = e^{-(d/\tau)^2}$
- probability initially remain high (the *shoulder*), reflecting an increased chance of detecting individuals closer to the observer
- practical advantages that we will discuss shortly (τ^2 is variance of the unfolded Normal distribution, $\tau^2/2$ is the variance of the Half-Normal distribution)

Negative Exponential and the Half-Normal are special cases of $g(d) = e^{-(d/\tau)^b}$ that have the parameter b [$b > 0$] affecting the shoulder

Hazard Rate

- two-parameter model: $g(d) = 1 - e^{-(d/\tau)^{-}b}$
- parameter b ($b > 0$) affecting the more pronounced and sharp shoulder

Distance functions



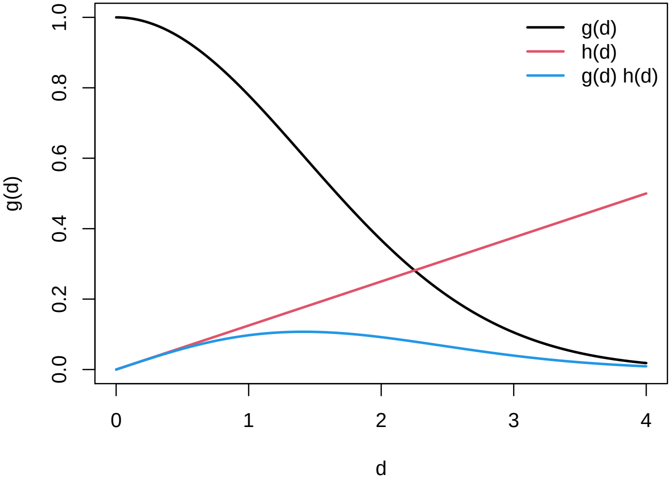
Distance sampling

The distribution of the *observed distances* is a product of detectability and the distribution of the individuals with respect to the point where the observer is located

- for point counts, area increases linearly with radial distance
- this implyes a triangular distribution:
  $h(d) = \pi 2d/A = \pi 2d/\pi r_{max}^2 = 2d/r_{max}^2$
- where A is a circular survey area with truncation distance r_{max}

The product $g(d)h(d)$ gives the density function of the observed distances.

Product



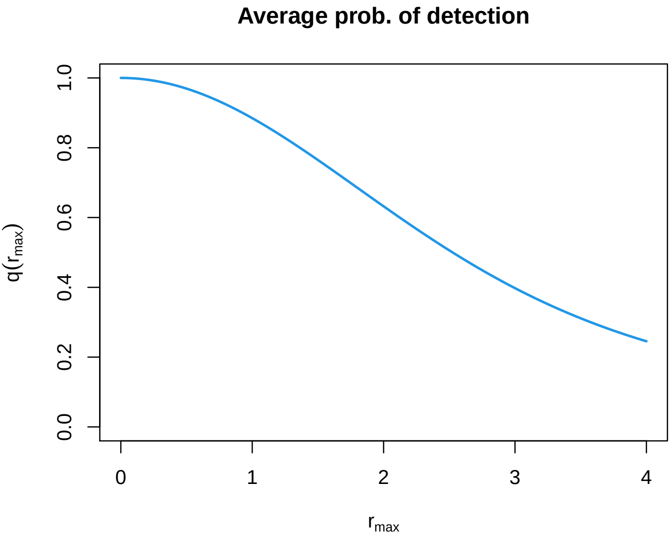
Average detection

Average probability of detecting individuals within a circle with truncation distance r_{max}

- we need to integrate over the product of $g(r)$ and $h(r)$
- $q(r_{max}) = \int_0^{r_{max}} g(r)h(r)dr$

This is the volume of *pie dough* cut at r_{max}, compared to the volume of the *cookie cutter* (πr_{max}^2)

Average detection: q



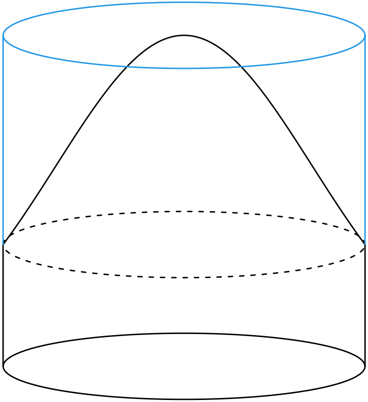
Half-Normal

For the Half-Normal detection function, the analytical solution for the average probability is

$$\frac{\pi\tau^2[1 - exp(-d^2/\tau^2)]}{\pi r_{max}^2}$$

where the denominator is a normalizing constant representing the volume of a cylinder of perfect detectability.

$q(r) = \text{black} \,/\, (\text{black} + \text{blue})$



Cell probabilities for Half-Normal

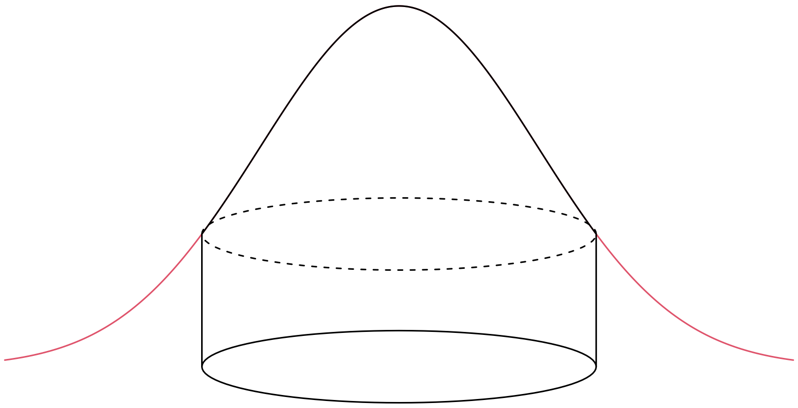
The cumulative density function for the Half-Normal distribution:

- $\pi(r) = 1 - e^{-(r/\tau)^2}$
- used to calculate cell probabilities for binned distance data
- the normalizing constant is the area of the integral: $\pi\tau^2$ (instead of πr_{max}^2)

It captures the proportion of the observed distances relative to the whole volume of the observed distance density.

In the pie analogy, this is the dough volume inside the cookie cutter, compared to the dough volume inside and outside of the cutter (that happens to be $\pi\tau^2$ for the Half-Normal)

$\pi(r) = \text{black} \;/\; (\text{black} + \text{red})$



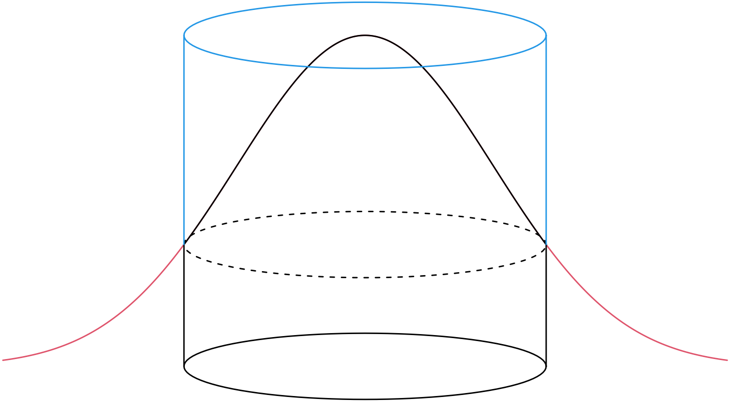
EDR

In case of the Half-Normal distance function, τ is the *effective detection radius* (EDR)

The effective detection radius is the distance from observer where the number of individuals missed within EDR (volume of 'air' in the cookie cutter above the dough) equals the number of individuals detected outside of EDR (dough volume outside the cookie cutter)

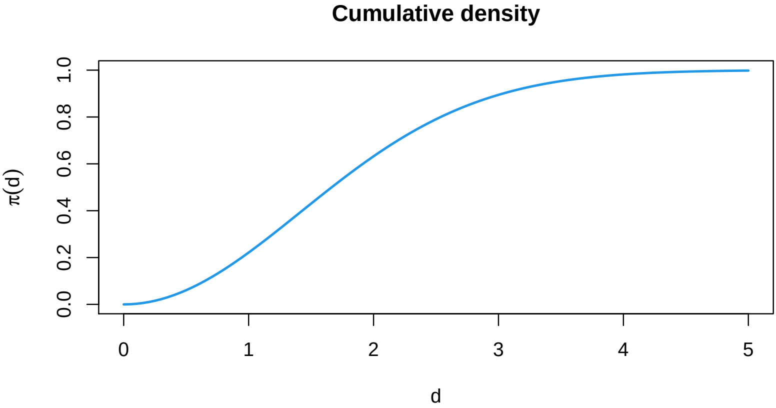
EDR is the radius r_e where $q(r_e) = \pi(r_e)$:

EDR: where blue = red

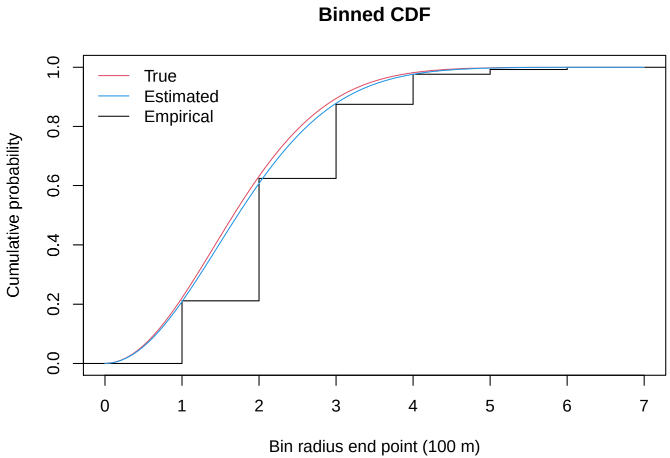


Estimation

The function $\pi(r)$ increases monotonically from 0 to 1



Binning


Binned CDF
True
Estimated
Empirical
Cumulative probability
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
7
Bin radius end point (100 m)

Traits can help

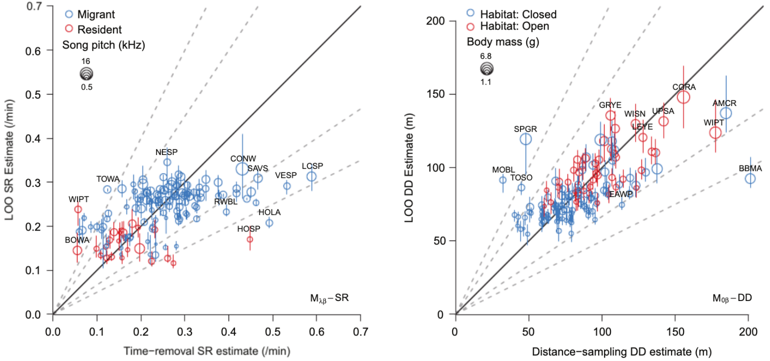


Figure 2. Leave-one-out (LOO) cross validation estimates vs empirically based estimates of singing rate (SR, model $M_{\lambda\beta}$; left) and detection distance (DD, model $M_{0\beta}$; right). Each circle represents a species (n= 141), symbol size and color vary according to inset legends. Error bars associated with each circle indicate conditional prediction intervals based on parametric bootstrap estimation. The solid diagonal line represents equality, and the grey guiding lines in order of increasing slope indicate 0.5×, 0.66×, 1.5×, and 2× ratios of the two estimates. Four-letter acronyms follow species' common names in Fig. 1, and are only included for outliers.