Introduction to Distance Sampling

Covariates in the detection function

Outline Solutions

1 Simulated whale data

An example of the sort of analysis you might have performed is given in the archived project file CovarWhaleSim-solutions.zip. If you sort by date created or analysis ID, you can see the order I set up the analyses in. I first tried simple half-normal and hazard rate models without covariates, and found that the half-normal model had a lower AIC. I then tried the MSTDO covariate and hour covariates separately (as non-factor covariates). The analysis with MSTDO had a much lower AIC, but the analysis with hour actually had a higher AIC than the analysis without covariates. I tried an analysis with both MSTDO and hour, but this had a higher AIC than MSTDO alone (Table 1). I concluded that the MSTDO covariate was important, but the hour covariate was not.

Although these data did not appear to need any truncation, I briefly confirmed that the same results were obtained with 10% truncation (analyses 7 and 8). Further analyses could look at the effect of adding adjustment terms to the detection function, although since no adjustments were selected with the half-normal without covariates it is likely that none will be required when the MSTDO covariate is used.

Table 1. AIC values for the candidate models.

Model	No truncation	10% truncation
HZ: simple model	125.32	82.46
HN: simple model	123.28	80.80
HN: with MSTDO	111.21	76.06
HN: with HOUR	125.03	82.15
HN: with HOUR + MSTDO	113.14	78.02

2 Analysis of golf tee data

With three covariates there are eight possible detection function models (including perpendicular distance only). The AIC from the CDS model was 311.1 and the lowest AIC I found was 304.3 which included sex as the only additional covariate. You may have found a different model. Table 2 is a summary of the results from the CDS analysis and an MCDS analysis with my best model. The component of variance due to the detection function fitted as a CDS was 64.3% and this reduced to 54.2% when sex was included in the detection function.

As part of an exploratory data analysis it is useful to analyse the data as a CDS but post stratify using the factor variables and fit separate components of the model for each factor level (as long as there are enough observations). The esw's for females (factor level = 0) and males (factor level = 1) are 1.61 metres (%CV=13.0) and 2.65 metres (%CV=10.3), respectively. The esw's for the exposure levels 0 and 1 are 2.41 (13.2) and 2.31 (10.0). The differences between males and females appear to be much larger than the difference between exposure levels indicating that sex would be the more useful covariate to include in the model. Notice how the abundance estimate for the CDS post-stratified by sex and the MCDS including sex are very similar, but the CV is smaller for the latter.

Table 2. Parameter estimates from CDS models and MCDS model which included sex only. CV's are given in parentheses

Parameter	True value	CDS	CDS post stratified by sex		MCDS
			Female (0)	Male (1)	
AIC		311.14	69.7	234.7	304.3
esw (m)		2.34 (7.9)	1.61 (13.0)	2.65 (10.3)	2.24 (6.4)
Ds (clusters per m^2)	0.15	0.13 (7.9)	0.05 (21.3)	0.08 (10.3)	0.13 (11.0)
E[S]	3.04	3.01 (5.9)	2.80 (13.7)	3.13 (6.5)	3.01 (5.9)
D (tees per m^2)	0.45	0.38 (9.9)	0.14 (18.9)	0.25 (12.2)	0.40 (8.8)
N	760	638 (9.9)	243 (18.9)	421 (12.2)	666 (8.8)
			664 (22.5)		

3 Analysis of dolphin sightings data

To obtain an overall impression of the data it is useful to fit a detection function histogram with many intervals (you may have problems fitting to the maximum number of 30, but 25 intervals should be OK). The spikes in the histogram suggest that the data has been rounded to zero and possibly other values. The q-q plot also indicates problems with the model at zero distances. To mitigate these problems, use the diagnostic tab to pool the data into a few intervals – 10 to 15 intervals work OK.

For the MCDS analysis, cluster size was fitted as a continuous variable, whereas, month, Beaufort, cue and search position were fitted as factor variables. Table 3 summarises the results. The number of adjustment terms allowed was limited to a maximum of two. In most cases a half normal function was chosen with either no, or one, adjustment term.

Table 3. Parameter estimates for the different models. Percentage CVs are given in parentheses. Note that CVs for the model containing cluster size are obtained by bootstrapping.

Parameter	CDS	Cluster size	Month	Beaufort	Cue	Search
AIC	3365.9	3359.5	3362.6	3366.9	3368.3	3339.8
esw (nm)	3.00 (4.5)	3.08 (1.9)	3.00 (1.9)	3.00 (1.9)	3.00 (1.9)	2.93 (2.3)
Ds (clusters per nm^2)	181 (4.5)	177	181 (1.9)	181 (1.9)	181 (1.9)	185 (2.3)
E[S]	507 (5.3)	460	529 (5.3)	507 (5.3)	495 (5.3)	589 (5.3)
D (animals per nm^2)	91965 (7.0)	81454	96009 (5.7)	91921 (5.6)	89729 (5.6)	109420 (5.8)

Based on the AIC, it seems as though the model including search method is best, however, there were warning messages about the detection function fitting and cluster size estimation. Before going on and looking at models which include two covariates, it is worth looking at the search model in more detail. The detection functions have very different scale parameters, for example, the detection function for search method 3 (using a helicopter) has a very wide shoulder and so the scale parameter is very large. This suggests that the observers were seeing everything out to 5 nm and so detection does not decrease with distance as it does with the other methods. One assumption of MCDS is that the perpendicular distance distributions of the covariate factor levels have the same shape. It may be worth refitting the model ignoring the observations made by the helicopter. Data can easily be selected/ignored using the Data filter | Data selection tab. The selection criteria will be of the form '[Search method] IN (0,2,5)'

This is a large dataset and so it is worth deciding on your final model before doing any bootstrapping to obtain variances.

4 Hawaiian Passerines

We provide no sample solution to these data; consult the Marques et al. (2007) reprint.