Design and Analysis of Distance Sampling Studies

Carl James Schwarz

Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, BC, Canada
cschwarz @ stat.sfu.ca

Part 3
Conventional Distance Sampling - Survey Design

Survey Design

Distance sampling - Survey Design

Do you need distance sampling?

- Not all animals can be detected in traditional strips or quadrats?
- No information needed on population parameters such as survival rates?
- Density is sensible but population abundance is ill defined.

Line vs. point sampling

- Line transect sampling uses all detection along route. In point transect, detections during travel between points is ignored.
- Line transect better at capturing animals close to line; in point transect, area searched close to point is small.
- Point transects good for fragmented habitats.
- Point transects good for multi-species (bird) surveys.

Distance sampling - Survey Design

Standard considerations

- What are your objectives? Density estimation?
- What is geographic range/scope of estimate?
- What subdomains do you need (e.g. density by sex? Density by habitat type?)
- Sampling time relative to animal movement.

Distance sampling - Survey Design

Good results require RRRs in survey design:

- Randomization makes the survey representative
  - Avoid subjective placement.
  - Systematic placement likely good a random placement (self randomization assumed) as long as systematic placement does not match feature in study area.
  - Run systematic transects perpendicular to known gradient in density.
  - All areas of study area should have equal probability of selection (at least within strata)
- Replication gives you adequate precision
  - Multiple short lines better than few longer lines.
  - 15+ replicate lines needed to get estimate of encounter rate variation.
- Stratification control for KNOWN noise.
  - Avoid too many strata (2 or 3 are usually sufficient).
  - Post-stratification is ok, but you may not have all information needed to rollup.

Distance sampling - Survey Design

Good results require attention to key assumptions

- $g(0) = 1$, i.e. animals detected on line/at point with certainty.
- Broad shoulder in detection function near 0.
- No movement of animals.
- Distance is determined accurately.
- Cluster size determined accurately.

Distance sampling - Survey Design

$$\widehat{D}_{Line} = \frac{n\widehat{f(0)}\widehat{E[S]}}{2L} \quad \widehat{D}_{Point} = \frac{nf'(0)\widehat{E[s]}}{2\pi k}$$

$$\widehat{var}(\widehat{D_{Line}}) = \widehat{D}^2 \times \left\{ \frac{\widehat{var}(n)}{n^2} + \frac{\widehat{var}(\widehat{f(0)})}{[\widehat{f(0)}]^2} + \frac{\widehat{var}(\widehat{E[s]})}{[\widehat{E[s]}]^2} \right\}$$

$$\widehat{var}(\widehat{D_{Point}}) = \widehat{D}^2 \times \left\{ \frac{\widehat{var}(n)}{n^2} + \frac{\widehat{var}(\widehat{f'(0)})}{[\widehat{f'(0)}]^2} + \frac{\widehat{var}(\widehat{E[s]})}{[\widehat{E[s]}]^2} \right\}$$

Distance sampling - Survey Design

- Variation in n (the encounter rate) is out of control of the experimenter but might be improved by stratification into high and low density areas. Aim for 15+ points/lines. Avoid pseudo-replication by sub-dividing lines.
- Variation in $f(0)$ and $f'(0)$ depends on number of detections. Improves as $\sqrt{detection}$.
  - 50+ distances for each separate detection function fit in line transect;
  - 80+ for point transects (to account for area searched effect).
  - w should be set large relative to average observed distance.
  - Don't concentrate on object far away as they will likely be truncated.
- Variation in cluster size is out of control of experimenter. Try using as a covariate or as a stratification variable.

Distance sampling - Survey Design

Absolute sample size is important and not relative sample size.

Distance sampling - Survey Design

Two rough rules of thumb based on a pilot survey.
Suppose that a target rse_t is desired (i.e. 95% ci is $\pm 2 rse_t$)

$$L_{required} \approx \frac{3}{[rse_t]^2} \frac{L_{pilot}}{n_{pilot}}$$

$$L_{required} \approx \frac{L_{pilot}[rse(\widehat{D}_{pilot})]^2}{[rse_t]^2}$$

Argus example,

- We want $rse_t = 0.10$.
- $L_{pilot} = 316$ km; $n = 57$
- $rse(\widehat{D}_{pilot}) = 0.21$

$$L_{required} \approx \frac{3}{[.10]^2} \frac{316}{57} = 1700 \text{ km}$$

$$L_{required} \approx \frac{316[0.21]^2}{[0.10]^2} = 1400 \text{ km}$$

Same equation for point transect replacing L by k.

Two rough rules of thumb based on a pilot survey.

- Additional effort is needed if measuring clusters. See the Intro book for more details.
- Use the above even if you tend to stratify as a rough first guess. Allocate among strata roughly proportional to $A_h\sqrt{D_h}$.
- Precision essentially improves as $\sqrt{effort}$.

What to do if you have little preliminary information?
A VERY, VERY, VERY, VERY rough LOWER bound for the effort needed is based on getting 100 distances based on a 50% average detectability on a strip half-width w when the density is D.
Solve for L such that

$$100 = 0.5 \times 2Lw \times D$$

Argus example,

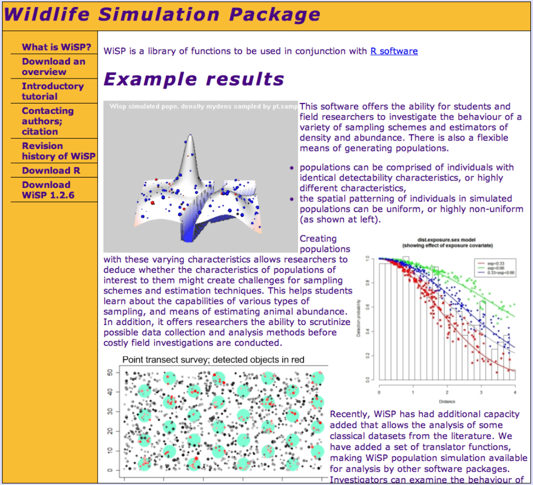
- $D \approx 0.025/\mathrm{ha} = .0000025/\mathrm{m}^2$
- $w \approx 100$ m.

This gives

$$L \approx \frac{100}{0.5 \times 2 \times 100\ \mathrm{m} \times .0000025/\mathrm{m}^2} = 800,000\ \mathrm{m} = 400\ \mathrm{km}$$

Distance sampling - Survey Design

What to do if you have little preliminary information?
Try *WiSP* package in *R*.


Wildlife Simulation Package

- What is WiSP?
- Download an overview
- Introductory tutorial
- Contacting authors; citation
- Revision history of WiSP
- Download R
- Download WiSP 1.2.6

WiSP is a library of functions to be used in conjunction with R software

Example results

Wisp simulated popn. density mydens sampled by pt.samp

This software offers the ability for students and field researchers to investigate the behaviour of a variety of sampling schemes and estimators of density and abundance. There is also a flexible means of generating populations.

- populations can be comprised of individuals with identical detectability characteristics, or highly different characteristics,
- the spatial patterning of individuals in simulated populations can be uniform, or highly non-uniform (as shown at left).

Creating populations with these varying characteristics allows researchers to deduce whether the characteristics of populations of interest to them might create challenges for sampling schemes and estimation techniques. This helps students learn about the capabilities of various types of sampling, and means of estimating animal abundance. In addition, it offers researchers the ability to scrutinize possible data collection and analysis methods before costly field investigations are conducted.

dist.exposure.sex model (showing effect of exposure covariate)

Point transect survey; detected objects in red

Recently, WiSP has had additional capacity added that allows the analysis of some classical datasets from the literature. We have added a set of translator functions, making WiSP population simulation available for analysis by other software packages. Investigators can examine the behaviour of

I have a simple script demonstrating how the package run in the WiSP Directory.

Distance sampling - Survey Design - Summary

- RRRs
- 15+ replicate lines; 40+ points
- 50+ distances for line transect; 80+ distances for point transect
- Do a pilot study to estimate how much effort will be needed in future.
- Use WiSP or other simulation features to estimate sample sizes in very difficult circumstances.