Design and Analysis of Distance Sampling Studies

Carl James Schwarz

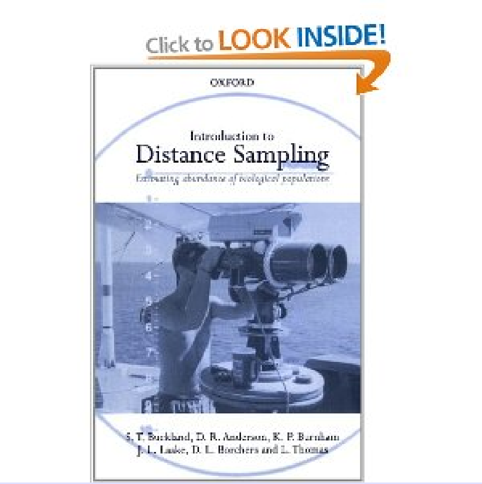
Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, BC, Canada
cschwarz @ stat.sfu.ca

Part 5

Whew!

Final Summary

Read this book for basics.



Distance Sampling - Final Summary

Use this software (unless you are an *R* geek.

`http://www.ruwpa.st-and.ac.uk/distance/`

Program Distance Home Page 13/11/2012 10:32 AM

Distance Home Page

Distance is supported by

Program Distance

- What is Distance?
  - Who uses Distance? - over 26,000 downloads!
- Download Distance
  - Distance 6.0 Release 2
- Distance Support, Updates and Extras
  - Report a Bug in Distance
- Sponsors of Distance
- Program Authors and Citation
  - Paper on program Distance (pdf) - now the default citation

Distance Sampling

- Distance Sampling Reference Material
  - Online preview of Introduction to Distance Sampling
  - Online preview of Advanced Distance Sampling
  - Non-technical introduction to distance sampling (pdf, short)
  - International Encyclopedia of Statistical Sciences article (pdf, short)
  - Encyclopedia of Environmetrics article (pdf, longer)
  - Pronghorn aerial survey manual
  - Extensive online bibliography
- Distance-sampling Email list
- Workshops and Meetings
  - NEW! Training workshops @ St Andrews, Aug 2012
  - Online abstracts from ISEC 2010 conference
  - Online abstracts from passive acoustic density estimation workshops 2009
  - Online abstracts from ISEC 2008 conference
  - Online abstracts from ICODS 2001 conference

Web site hosted by
Research Unit for Wildlife Population Assessment

Comments and suggestions to
distance@mcs.st-and.ac.uk

Last modified December 8, 2011

Distance Sampling - Final Summary

When to use distance sampling?

- Low density of objects that can be rapidly visually surveyed (e.g. from airplane)
- Only density/abundance is of interest
- Objects are relative immobile relative to observer.
- Detection falls a function of distance.

Distance Sampling - Final Summary

When NOT to use distance sampling:

- Need information on survival rates and other population parameters
- Need to follow individual animals.
- Animals are hidden (multipliers?); animals move very fast
- Animals avoid/attracted to observer.

Distance Sampling - Final Summary

Types of Distance Sampling (not a complete list)

- Line transect sampling - this course
- Point sampling - this course
- Cue counting - see book
- Double observer transect sampling - this course
- Spatial density sampling - next course

Distance Sampling - Final Summary - Protocol

Survey Protocol: Estimate **Density** and/or **Abundance**

- Move along a transect and measure objects seen from the transect
  Stay at a point and measure distance of objects seen from the point.
- Detection probability on the transect line/point is certain (probability $= 1$) unless you use MRDS.
- Detection probability declines with distance from the transect/point

Distance Sampling - Final Summary - Assumptions

0. No assumptions made about distribution of animals on landscape!

- D describes the density over the landscape.
- Not necessary that animals distribute themselves independently, i.e. aggregated distributions are allowed.
- RANDOM placement of transects in large study areas implies uniform distribution of perpendicular distance from transect line.
- In small areas, strip width may go off site, leading to potential bias.

Distance Sampling - Final Summary - Assumptions

1. Animals on transect line are detected with probability $= 1$ unless you use MRDS.

- Blind strip along line (aerial surveys, animals hide) dealt with by left truncation.
- If $g(0) < 1$ then estimates of D are biased downwards.
  - If a correction factor is available, specify a multiplier, (e.g. only males visible) and sex ratio is 1:1.
  - Multiple observer/ platform methods.
- Caution about "guarding the centerline" can lead to detection functions that are hard to model.
  - High detection rates close to transect line, rapid fall off, then usual decline is hard to model.

Distance Sampling - Final Summary - Assumptions

2. Animals detected at initial location
   - Movement of animal independent of observer causes little bias if movement is slow
     - CAUTION: Movement could cause an object to be counted twice (herding).
     - CAUTION: Fast moving ($> 1/2$ of observer speed) animals lead to severe bias.
   - Counting of same animal on different transects is ok, as long as movement is random.
     - Animals moving to other transect "balance" animals moving away from other transect.
   - CAUTION if animals are attracted or frightened near the transect line this can lead to serious bias.
     - E.g. animals being herded as observer moves along transect line.
     - Histogram of distances often show hole along transect line.

Distance Sampling - Final Summary - Assumptions

3. Measurements of distance from transect line are exact.
   - Even random errors introduce bias because "inflation" by detection probability is larger further from the transect line.
   - Visual estimates of distance are NOT very good, i.e. observers tend to underestimate distances.
   - Beware of heaping, i.e. lots of measurements at 100 m, but fewer at 99 m and 101 m.
   - Try using interval data (binned) (e.g. distances estimated by airplane struts).

4. Beware of outliers.
    - Objects far away have large influence on estimates because of weighting by 1/prob of detection.
    - Check histograms and truncate.
        - Truncation introduces no bias.

Distance Sampling - Final Summary - Assumptions

5. Objects are detected independently of each other.
    - Clusters of animals violate this assumption, but can be dealt with if cluster size is recorded.
    - Pair-bonds of birds where some members of cluster may be hidden?
    - Ill-defined clusters?
    - Clusters that are too large to count?

Distance Sampling - Final Summary - Assumptions

6. Detection function is not spiked at 0.
    - Methods work best if there is shoulder to the detection function.
    - Negative exponential (with no series adjustment) is a last resort!

Distance sampling - Final Summary - Common Problems

Single Transect Line

- Not possible to use empirical estimate of variance.
- Valid if want to make inference only about area sampled (i.e. along the transect line), but this is doubtful.
- Assume distribution of observations is Poisson with an over dispersion factor of about 3.
- Refer to DISTANCE manual.

Distance sampling - Final Summary - Common Problems

Missing Data - I

- Missing Distances
  - Nothing entered into DISTANCE.
  - No bias introduced assuming not on line, but loss of precision.
  - If MCAR and on the line, estimate P and use as multiplier
- Line with no observations
  - Do NOT discard the transect.
  - Enter line with length and NO observations.
- Line could not be run due to weather or other factors.
  - Assume missing line is MCAR and do NOT enter in DISTANCE
  - If only part of line can be run, enter the observered portion into DISTANCE.

Missing Data - II

- Missing Cluster Size
  - By design, i.e. record all clusters, but only count cluster size in every 3rd cluster.
    Enter -1 for cluster size;
    Objects are used to estimate cluster density, but non-zero sizes used to estimate mean cluster size which are then multiplied together.
  - Empty objects, i.e. look for dens and count number of occupants of dens.
    (a) Discard empty dens; use non-zero values; biased estimates of average den size, but fewer dens so ok
    (b) Use empty dens; but cannot adjust for size-biased sampling (not a problem in this example).
- Missing Survey effort (length of transect) – Go home early, nothing can be done.
- Missing Study area size – you can estimate D but not N.

Distance sampling - Final Summary - Analysis

Collect data:

- Use good survey design to collect data (see planning later in the course).
- Enter data into spreadsheet
    - Group items together (e.g. observations from the same transect together)
    - Think of stratification variables (pre- and post-) and create
    - Include covariates (see later in this course)
    - Don't just drop missing data
        - Transects with NO observations have blanks for distance from transect line
        - Cluster size 0, e.g. detect dens from transect line that is empty; enter 0 for size

Distance sampling - Final Summary - Analysis

Preliminary exploration of data:

- Don't plunge into analysis without exploration of data.
- Look for
  - Outliers (large distances have low probability of detection) have large inflation.
  - Heaping (distances are 100 m but never 99 m or 101 m).
  - Shape of distance distribution - should have shoulder, monotonic decline, going toward zero.
  - Hole at zero - indicates animal avoidance etc.
- Consider converting "exact distances" to intervals if problems exist such as heaping.

Import data into DISTANCE

- Save as *.csv or other text-delimited file.
- Separate folder for each project.
- Keep distance files together.
- Do regular backups.

Distance sampling - Final Summary - Analysis

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

- Right truncate at $g_{Line}(w) \approx 0.15$ $g_{Point}(w) \approx 0.10$
- Truncate at fixed width. [IMPORTANT.]
- Detection function
  - Only contribution to the likelihood which has implications later
  - Shoulder at 0; no spike at 0; left truncation if needed.
  - Monotonic decline; group distances if measurement error.
  - Decline to 0 as distance increases.
  - Many key function + series give similar fit but avoid neg. exp.
- Deal with cluster size via modeling, stratification, or covariate.
- Multipliers for known effects (e.g. seeing fraction of population, dung counts, etc.)
- Stratification/covariates can improve precision but don't go overboard.
- AIC for model selection but cannot change data filter among models being compared.
  - No simple way to model average, but see DISTANCE manual.

Distance Sampling Analysis - Super Summary

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

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

- Some components of variation you have little control over and are the same regardless of effort.
- Variance components are estimated by:
  - Empirical variance methods over replicate lines/points are robust for $var(n)$, but with small sample sizes, may be forced to specify Poisson-process.
  - MLE theory gives $\widehat{var}(\widehat{f(0)})$ and $\widehat{var}(\widehat{f'(0)})$
  - Empirical variance methods for $\widehat{var}(\widehat{E[s]})$

Distance sampling - Final Summary - Analysis

Review the output:

- Check model fit using QQ plots; KS/CM tests; χ^2 tests, but don't rely exclusively on these
  - Horizontal lines in QQ plots are artifacts of heaping and rounding.
  - Small expected counts in χ^2 test can inflate test-statistic and lead to false-poor fit. Look at pooling small cell counts.
- See if estimates are reasonable.
- Rank models of SAME data using AIC. Model averaging must be done outside of DISTANCE.
  Cannot use AIC for model comparison with different data filters.

Distance sampling - Final Summary - MRDS

Deals with $g(0) \neq 1$.
Survey Protocol

- Independent configuration.
- Trial configuration.
- Removal configuration (not yet implemented in MRDS).

Data structure:

- The usual + object identifier +
- Field for Observer (limit currently is 2) +
  Detected (1=yes, 0=no)

Group the data together in the usual way. Import in the usual way.

Distance sampling - Final Summary - MRDS

Model Building.

- Start with small models and build up to more complex models
  - Full independence vs. Point independence
  - Few covariates vs. many covariates
  - Complex models lead to cases where categories of animals seen by one observer are never seen by the other observer leading to an infinite population size!
- DS (HN or HR, but no series); MR (logistic regression)
- Use Linear Model syntax, e.g. distance + sex + distance:sex
- AIC for model comparison over all options within same data filter

R package has more features, but less easy to use.

Distance sampling - Final Summary

More advanced topics

- Density Surface Modelling (DSM) which allows you to see how the density varies over a landscape
- integration of DS with GIS esp with survey design
- 3-D distance sampling.
- Acoustic and other cue counting methods.
- Trapping Webs.
- Using *R* for more advanced features.

Distance sampling - Final Summary

cschwarz@stat.sfu.ca