

cite Buckland et al. 2001 Ariefiandy et al. 2013
cite* 2001 2013
parencite (Buckland et al. 2001) (Ariefiandy et al. 2013)
parencite* (2001) (2013)
textcite Buckland et al. (2001) Ariefiandy et al. (2013)
smartcite - does the super script/foot note thing.

Master's Report - Variable Circular Plots: Station Placement and the Independence Assumption

Matt Edwards

26 May 2014

1 Introduction

Estimating population density of plants or animals is a persistent issue in the ecological sciences. Distance sampling, in the form of line and point transects (alternatively, Variable Circular Plots (VCP)), is a common estimation method. They are useful when populations are too small to be seen from the air (birds, rodents) or may be masked by the canopy (ground plants.)

Distance sampling has been around since the early 1900's. In the 1970's, more rigorous scientific methods were applied, including papers by Ramsey and Scott (1979; 1981) which augmented work done by Emlen in 1971. Burnham, Anderson, and Jeffrey L. Laake 1980 is a monograph on the theory and application of line transect sampling. Non-parametric methods have been examined by Quang (1993) and Mack and Quang (1998). Distance sampling continues to be used and modified, combined with mark-recapture methods (J. L. Laake et al. 2011), extended for use with underwater acoustics for estimates of krill populations (Cox et al. 2011), and a study by Marcus Rowcliffe et al. (2011) combines distance sampling with camera traps to estimate populations.

Buckland et al. (2001) has a more in-depth discussion of the history of distance sampling.

Recent studies that use distance sampling include Norwegian bird populations (Pederson et al. 2012), Komodo dragon prey estimates (Ariefiandy et al. 2013), Serengeti carnivore abundance (Durant et al. 2011), and estimates for butterfly populations (Isaac et al. 2011).

With Line Transects, a (straight) line is randomly placed in the study area, then walked by the observer, with any objects of interest noted and the distance perpendicular to the transect recorded. With point transects, the observer stands in one spot, recording the direct line distance from herself to the object of interest as if it was projected on the ground. This results in data similar to the line transect, but with more safety for the observer as they do not have to watch for (moving) wildlife and watch where they are walking Ramsey and J Michael Scott 1979. VCP have the added benefit of allowing the wildlife to settle. Animals may react to an observer moving through the area by hiding or moving away. If the observer is standing still, and has allowed a "cooling" period for things to return to normal, animal behavior may be more natural and less influenced by the presence of the observer.

Observations of objects can take the form of visual sightings or detection of auditory cues, or one confirmed by the other, as long as an accurate distance can be measured.

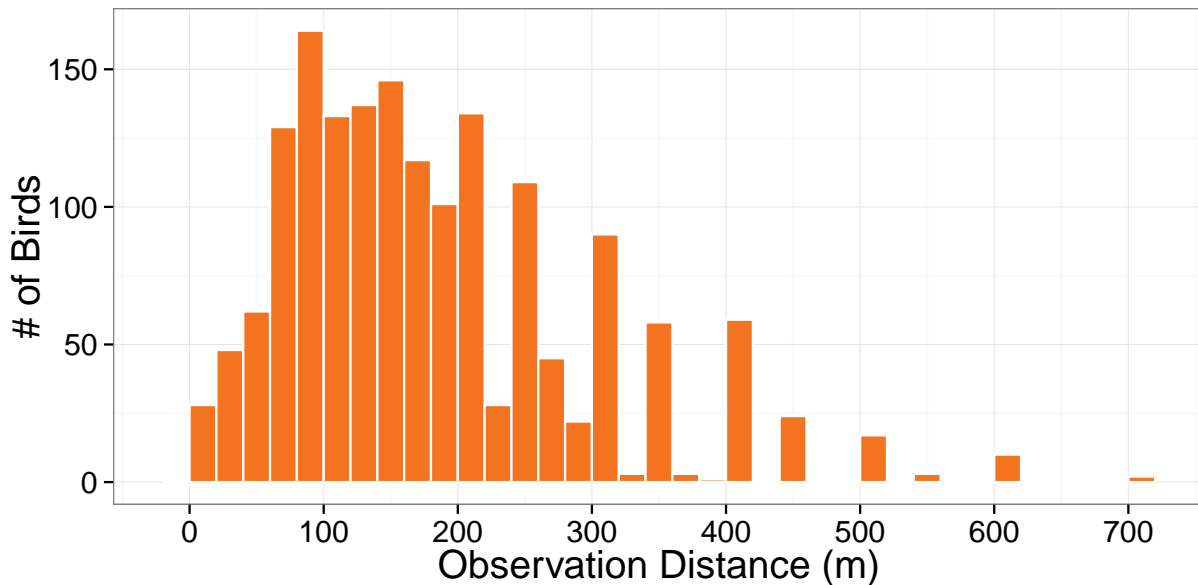


Figure 1: Aggregated Collared Kingfisher Observation Distances, Rota & Tinian, 20 m bins: Engbring, Ramsey, and Wildman 1986

Buckland et al. (2001, p. 6) draws a distinction between point transects, where only the area immediately surrounding the point is surveyed (area of estimated perfect detectability), and Variable Circular Plots, where you census the area πw^2 , where w is the furthest distance at which you can observe the object of interest. However, in Buckland (2006) and Quang (1993) the terms are used interchangeably. They will be used interchangeably here also, with the understanding this paper addresses the second type, where an area of $\pi * w^2$ is censused.

1.1 Micronesian Bird Data

The Micronesian Forest Bird Survey Engbring, Ramsey, and Wildman 1986 is a work using Variable Circular Plots (VCPs) placed using a random-systematic transect design to monitor the density of several bird species in the Micronesian islands.

Each of the five islands was divided into regions. Initial transects were given random positions and angles, with subsequent transects placed parallel, 2km apart (the “random-systematic” method). Stations were placed every 150m along the transects. Two observers, standing 20 m apart, would wait several minutes for the wildlife to quiet down after their arrival, and then spend 8 minutes noting any birds they could observe either visually or aurally. Distance and species were noted, as well as other covariates such as foliage density, and weather conditions.

Figure 1 illustrates the distance data for the Collared Kingfisher. This data represents the aggre-

gated data from two islands, Tinian and Rota, and all four observers. We can observe several things in this plot:

First, there are very low counts of birds immediately around the station, a peak at around 100 m, a fairly level section from 100–200 m, and a drop off after 200 m. This would seem to indicate that the presence of the observers caused either movement away from the station (the crest around 100m) or alternatively a suppression of vocalizations and/or movement directly around the station.

Secondly, the further away from the station, the distance observations were heaped into convenient distances. With smaller bin widths, this is also evident in the 100–200 m range as well.

Lastly, note that the detection distances are high between 100–200 m, and recall that the stations were placed 150 m apart. Since the stations on a transect were (generally) visited in sequence, this means it is possible a bird could have been observed from more than one station.

1.1.1 Independence

Both Ramsey and J Michael Scott (1979) and Buckland (1987) discuss the assumption that the VCPs are placed randomly through the study area. This would seem to implicitly state that their observational ranges might by chance overlap, meaning a single object of interest could be observed from more than one station. Thompson (2012, p. 240) likewise indicates random placement of line transects which would seem to allow for a potential overlap of observation zones.

Buckland et al. (2001) discuss the random placement of both line and point transects, without addressing the potential for sightings to overlap. Reynolds, J. Michael Scott, and Nussbaum (1980) state that the possibility of observing the same bird from two stations should be avoided. Buckland (2006) says simulations showed that analysis was robust to violations of independence (meaning overlap, or multiple observations on same object).

Buckland et al. (2001, p. 233) discuss that each line or point must be “randomly and independently located,” with “an equal probability of selection” for all portions of the study area.

Buckland et al. (2001, p. 235) states “Transects are normally spaced at a sufficient distance to avoid detecting an object from two neighboring transects, although this is not usually critical unless sampling a line changes the animal distribution at neighboring, as yet unsampled lines.”

They do state that for random-systematic transects, the methodology used by Engbring, Ramsey, and Wildman (1986), that each transect line should be treated and analyzed as a cluster.

In discussing line transects, Barry and Welsh (2001) treat it as a two-stage sample, with non-unique clusters (each study object can be in more than one cluster) having an equal probability of selection. This would seem to (but does not explicitly) endorse the possibility of overlapping VCPs.

One of the assumptions of distance sampling is that the objects being studied are distributed homogeneously with some uniform density, D , throughout the space Ramsey and J Michael Scott (1981). (If the data is clustered we have other problems, not addressed here.) If the uniform distribution is the case, then it would make sense that it shouldn't matter if the observations overlap, as long as the act of observing at one station does not affect the observations at the neighboring station. This echoes Buckland's statement a few paragraphs previous.

Thompson (2012, p. 244) mentions (page 244) that the systematic selection of transects (random position of first, with additional transects a fixed distance apart) does not affect the approximate unbiasedness of the density estimator, but will have an effect on the approximate unbiasedness of the variance estimator.

However, it is clear the literature is divided and unclear on whether or not our observation areas should overlap.

1.2 Assumptions

Buckland et al. (2001) lists three primary assumptions:

1. Objects at the line or point are detected with certainty $g(0) = 1$
2. Objects are detected at their initial location
3. Measurements are exact

There are additional assumptions listed by both Buckland and Ramsey and J Michael Scott (1979) and summarized by Quang (1993):

- iv Sightings of animals are independent events
- v at least one animal is detected $n > 0$
- vi detectability is the same for all animals at distance r and in all directions.
- vii no animal can be detected beyond a finite distance $w : g(r) = 0$ for $r > w$
- viii the function $g(r)$ has at least 4 bounded derivatives on $[0, w)$, and $g'(0) = 0$
- ix the stations are chosen at random in the study area

Ramsey 1979 Uniform distribution, random process distribution, independent distribution,

Not explicitly stated, is that we are assuming the objects of interest are homogeneously distributed throughout the region being surveyed. REFERENCE

1.3 Broad Strokes

In general, an estimate of population density is a count of objects observed divided by the area in which they were observed times the probability of observing the object. (Ramsey, Scott 1979)

Equation 1: $D = E[n]/(A_r)(P_r)$

In practice, the expected value of n is taken to be n , the number of objects observed. The estimation of the denominator is where things get tricky.

If we combine $(A_r)(P_r)$ into an effective area we get:

Equation 2: $A_R = 2\pi \int_0^R g(y) y dy$

$G(r)$ is the detectability function, and represents the probability of detection at distance r . It is often scaled so that $g(0) = 1$. (In terms of area of a circle, it is obvious that the *integral* $g(y) y dy$ represents r^2 in the $2\pi r^2$ equation.)

Much of the uncertainty in VCP estimation comes from how the effective area is estimated. $g(y)$ is often modeled as a parametric class (exponential, half-normal, hazard function) which is then scaled so that $g(0) = 1$. In addition to guessing which function to use, we must also estimate the parameters.

It is plausible to think that as objects are further away from the observer, they are more difficult to detect, and therefore, we would have a greater probability of missing observations that were further away. So, in calculating $g(y)$, do we assume perfect detection out to the furthest point of observation? Or do we truncate our observations at some distance closer to the observer? Ramsey & Scott (1979) refer to these as the effective radius (ρ) and the basal radius (r) respectively.

The problem lies in that in practice, we don't know where the basal radius is, and it must be estimated from our data.

1.4 Properties

1.4.1 Bias & Consistency

Buckland et al. (2001) and Ramsey & Scott (1970) both stated that if our detectability estimate $g(z) = 1$ out to r , then D is unbiased.

Barry and Walsh (2001) compared design-based and model-based estimates of density and variance. (Their focus was line transects, but they state it is analogous to point transects and the conclusions should carry across.) They concluded that model-based estimates were unbiased if and only if the assumptions of objects being independent and uniformly located held. They noted that in practice, this is probably not a feasible assumption. For design-based inference, they caution about both the density and variance estimates.

Roeder, Dennis, & Garton (1987) use simulations to demonstrate that no method is completely applicable in 100% of situations, and will go sideways at some point, depending on the specific conditions.

The overall conclusion is that, in theory, density estimates using VCPs are unbiased and (reasonably) consistent, but that the assumptions necessary for the theory to hold are untenable.

As stated in the section on choosing $g(r)$, it will vary depending on the specific situation, and it is up to the scientist and statistician to understand the data and choose the best approach.

2 Simulation

As seen in the previous section, the issue of independence of the point transects does not have a consensus in the literature. If independence is required for proper estimation, then the random-systematic placement method should not be used.

Through simulation, we can explore what happens under three different VCP placement schemes, and compare the resulting density estimates to their known values to evaluate performance.

2.1 VCP Arrangement

Simulation will address three options for VCP arrangement:

- Structured: VCPs placed so that observation distances do not overlap
- Random: VCPs placed completely at random
- Random-Systematic: Two transects of 18 stations each, placed according to Ramsey et al's design: 150 m between stations, 2 km distance between parallel transects.

Figures 2 and 3 illustrate the first two arrangements. Density in these images is based on the 20 birds per km^2 of the Palie region on Rota. The surveyed area of the Palie region totaled 9.41 km^2 during the original survey, and contained two transects of 16 and 17 stations each, 33 stations total (Ramsey et al 1982).

As seen in Figure 1, there is very little detection happening beyond 500 m. There are 15 observations beyond 500 m, which is 0.9% of the data.

Treating 500 m as our truncation distance w , the study area in the simulation has been raised to 36 km^2 to allow for stations where the observation area did not overlap. An area of 49 km^2 is generated with the given density, then truncated to 36 km^2 to prevent any edge effect from the random generation of points. The stations are placed a minimum of 500 m (0.5 on the graph) away from the edges to prevent border truncation.

For the random-systematic, an angle was chosen between 0 and π . A transect of 18 stations 150 m apart was constructed, and a second transect of 18 stations constructed running parallel. The transects were then randomly placed within the graph, no closer to the edge than 500 km (0.5 on graph.) Figures 4 and 5 are two examples.

3 Simulation Setup

3.1 $g(0)$

3.1.1 Half Normal

For a maximum possible observation distance of 0.500 (500 m), the parameter for the half-normal detection function is $\theta = 8.77$, with a scale parameter $\delta = 0.179$.

The half-normal parameter θ is related to the standard deviation of a standard normal distribution by the relationship:

$$\theta = \frac{\sqrt{\pi/2}}{\sigma}$$

If we estimate σ by $w/3.5$ (with w being the maximum distance at which we might observe an object, and 3.5 being the standard deviation point where $P(X > 3.5) < 0.001$) then theta is:

$$\sigma = w/3.5 = 0.1429$$

$$\theta = \frac{\sqrt{\pi/2}}{0.1429} = 8.7732$$

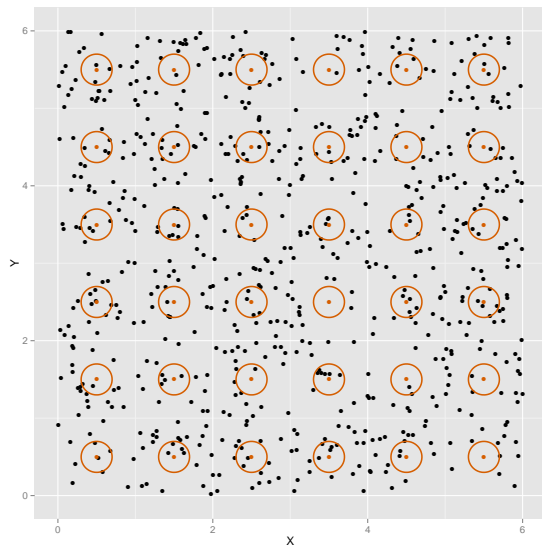
For a half-normal with parameter $\theta = 8.7732$, $f(0) = 5.5852$ so to scale the value to 1:

$$\delta_{HN} = 1/5.5852 = 0.1790$$

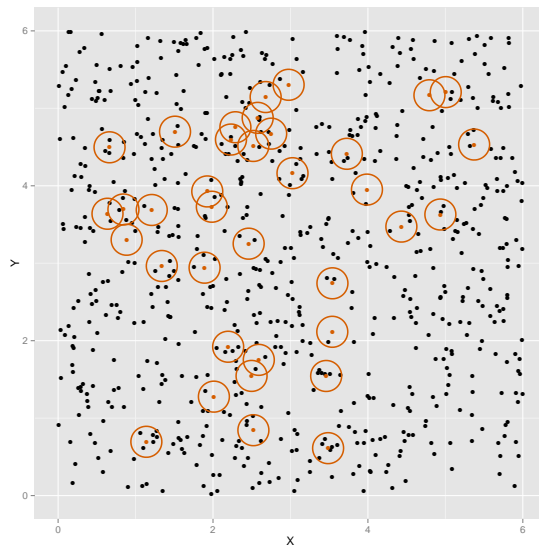
3.1.2 Empirical Observations

We can see in Figure 4 that above 100 m, observations were commonly heaped into 10 or 25 m distances, with the latter being more heavily used. Above 200 m, 50 m heaps were common. To estimate a density function, I used a loess regression with a span of 0.2 on the counts from 5 meter bins of the empirical data. This is the blue line in Figure 4.

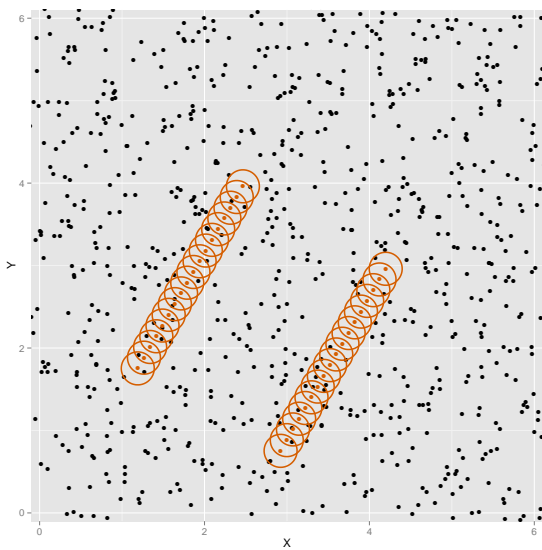
Using the `hist()` function in R, the 1982 observation distances were broken into counts by 5 m groups. Using the bin midpoints, the density estimate was multiplied by 5 to get the density value for that bin. A LOESS regression was run with the midpoint as x and $y = density * 5$, using the `loess()` package in R.



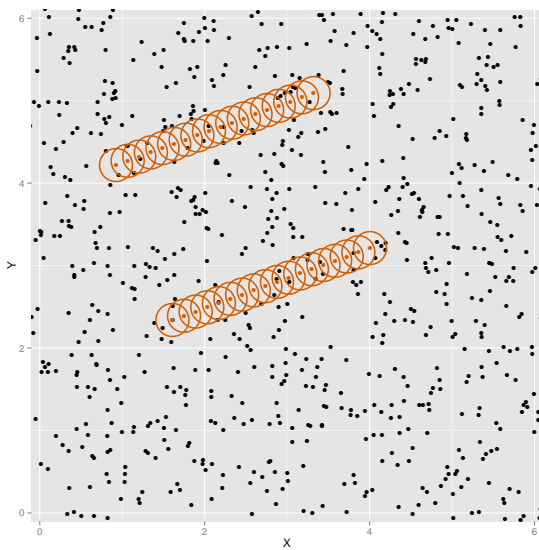
(a) Structured Layout



(b) Totally Random Layout



(c) A Random-Systematic Layout



(d) Another Random-Systematic Layout

Figure 2: VCP Layout options. On graph, 1 unit = 1 km. Circles represent 200 m mark.

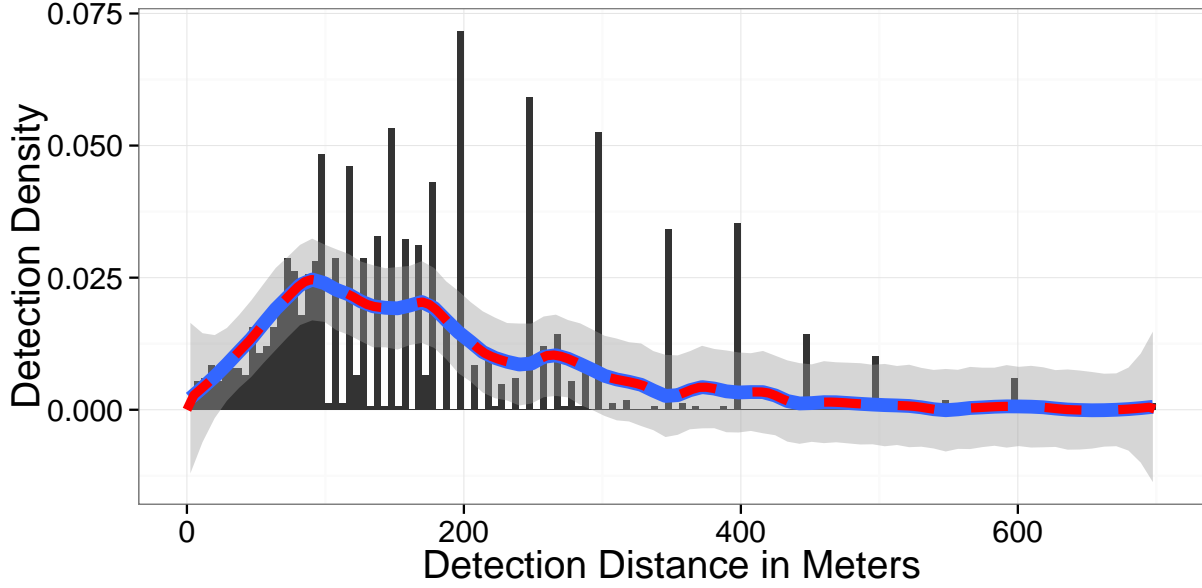


Figure 4: Empirical Observation Distances from Original data in meters, with LOESS regression lines. Bin width=5 m

This gives us an object that can be used with `predict()` to get density estimates for specific distances. The values generated by feeding our midpoints into our prediction function are represented by the red dashed line in Figure 4.

Following the scaling of the highest point to 1, as was done with the half normal curve, where the highest point was $g(0)$, we also find δ for this empirical detection probability. A sequence of x values was generated covering the range from 0–500 m. This was plugged into the `predict()` function using our LOESS object. The highest density value was then scaled to be 1:

$$\delta_{EMP} = \frac{1}{\max(\text{predicted})} = 40.5780$$

I know I read about how to deal with the random-systematic method, but now, of course, can't find it. With Line transect sampling, they essentially treat the data as if it came from one long transect, so that's what I'm going to do until I can find another reference.

4 Results

4.1 Half-normal vs. Empirical $g(x)$

Figure 5 shows the results of 1000 simulations, each analyzed with Empirical and Half-normal detection function, $g(x)$, and using each of the three layouts, Structured, Random, and Transect.

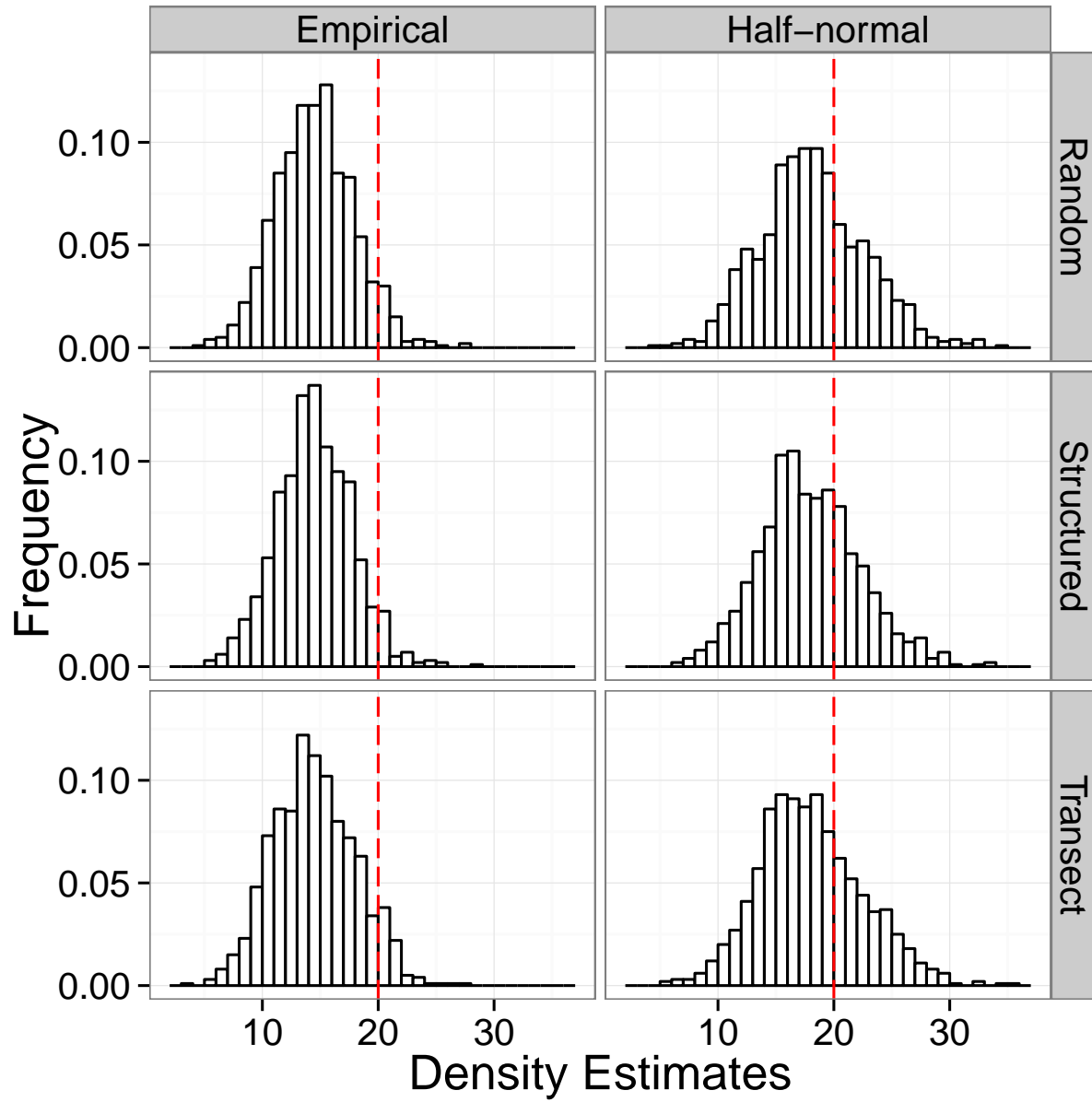


Figure 5: 1000 Simulations, Empirical vs Half-normal Detection function, for three different layouts.

$g(x)$	Layout	Mean	Std Dev	LB 95	UB 95	25th	Median	75th
Empirical	Random	14.52	3.35	7.97	21.08	12.17	14.46	16.74
Empirical	Structured	14.50	3.22	8.19	20.81	12.38	14.45	16.62
Empirical	Transect	14.46	3.53	7.55	21.37	11.85	14.36	16.88
Half-normal	Random	18.17	4.50	9.34	27.00	15.24	17.94	21.00
Half-normal	Structured	17.94	4.29	9.54	26.35	15.19	17.67	20.78
Half-normal	Transect	18.04	4.49	9.23	26.84	14.90	17.75	20.74

Table 1: Empirical vs. Half-normal detection function, across all three layouts. 1000 simulations.

All \hat{D} were done using the kernel method with a normal kernel (Quang 1993).

For each simulation, a single layout of “objects” is generated, and that map is analyzed by each of the 6 combinations of VCP layout and $g(x)$.

As we would expect, the density estimates using $g(x)$ based on the Empirical distance observations is biased low. We also see bias in the estimates using the half-normal function for $g(x)$, but it is much less severe than in the Empirical data.

Addressing our primary question of interest, does the overlap of the VCP affect the population density estimate due to violation of independence, the answer would appear to be no. Looking at the simulations with the Half-normal detection function, the histograms are fairly similar for the three layouts.

T0 D0: along with d.hat, get the variance estimate for d.hat, and return T/F if the true value was contained in the interval.

T0 D0: look at 3 moved categories for the 3 layouts

5 The Problem with Circles

To test the effectiveness of my movement algorithm, I compared the detection distances from the data where the objects moved, to the detection distances from the same data, assuming they stayed still. The movement data, (left panel, Figure #) is squashed more towards 0.1 Km than the still data (right panel). But there’s a problem. The still data doesn’t show the nice shoulder at 0 that the literature instructs us to expect.

T0 D0: Images

Figure 6 shows the distances from 36 randomly placed VCPs to the objects within the max observation distance. As we can see, there are simply fewer objects close to the observer, which increases as the distance moves away. (Rephrase) This is an artifact of the VCP being a circle.

T0 D0: Same graph for Line transect

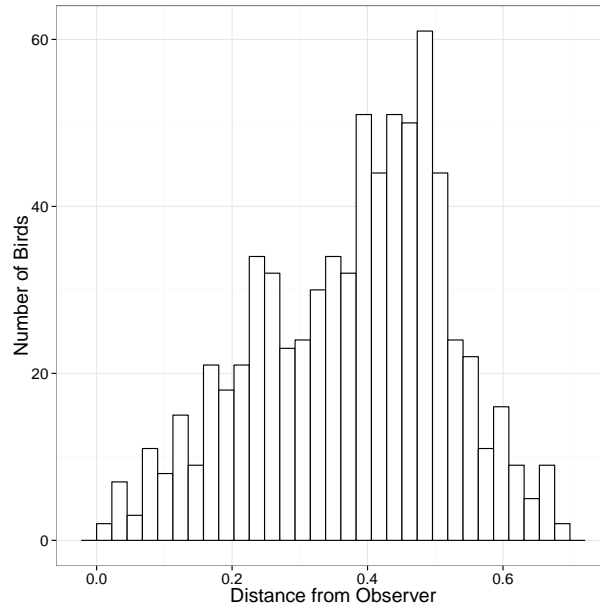


Figure 6: Simulated bird distances from center point of 36 randomly placed VCPs.

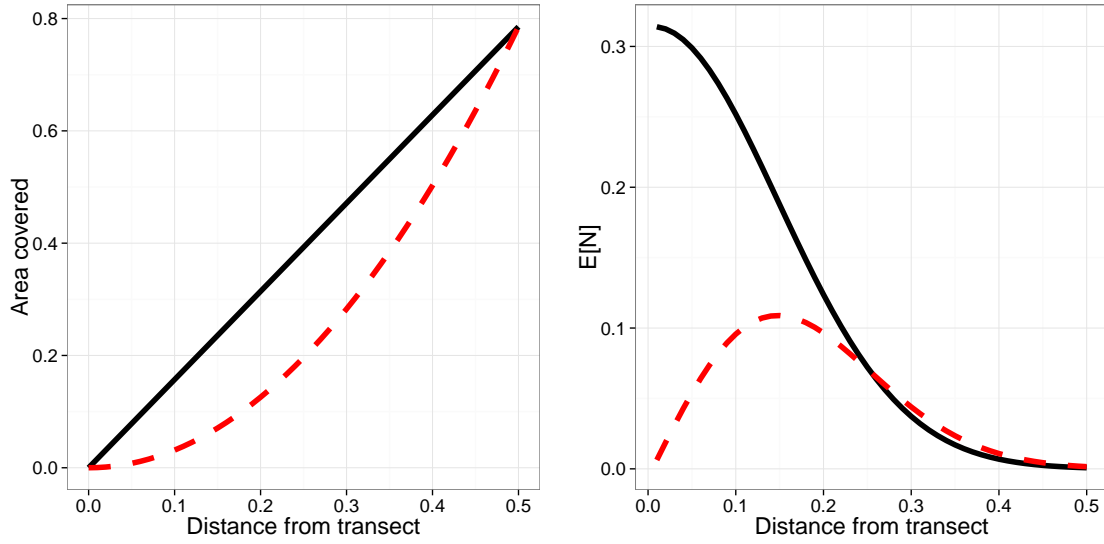
With a line transect, as we move away from the line, the distance "surveyed" increases linearly. With a point transect, as we move away from the point, the distance surveyed increases exponentially.

T0 D0: Equations showing what happens when you compare distance at r vs distance at $r+x$

Thought - is this why we use $h(0)$ for the point transects? need to verify, $h(r) = f'(r)$?? budkland 1987 paper has $h(r) = f(r)/r$, need to check teh book.

5.1 Notation

This notation is borrowed from Quang (1993) following Buckland (1987)



(a) Area of rectangle vs circle at same distance (b) Estimated detection numbers for rectangle (solid) vs circle (dashed)

Figure 7: Squares (Line Transect, solid) vs. Circles (Point Transect, dashed)

- t : Number of plots surveyed
- D : Common density of animals in all plots
- m_j : Number of detected animals in the j th plot
- \bar{m} : Average number detected per plot
- μ : Expected number of animals in a plot, $\mu = E(m_j)$
- n : Aggregate number of animals detected in all plots, $n = \sum_{j=1}^t m_j$
- λ : Expected total number of detected animals in all plots, $\lambda = E(n) = t\mu$
- R_i : Distance from station to the i th detected animal
- R_1, \dots, R_n : Detected distances pooled from all plots
- s : Sample standard deviation of R_1, \dots, R_n
- $f(r)$: Common probability density function (pdf) of the detected distances, R_1, \dots, R_n
- h : Bandwidth used in kernel estimation

For a typical individual plot:

- N : True number of animals in the plot (unknown)
- m : Number of detected animals in the plot
- $g(r)$: Radial detection function, $g(r) = \Pr(\text{Animal is detected} \mid \text{Animal is at distance } r)$
- w : Distance beyond which no animals can be detected (= radius of the plot)
- p : Mean detection probability within the plot
- B : $B = 2/(pw^2)$

6 Bibliography

References

- Ariefiandy, A., D. Purwandana, G. Coulson, DM Forsyth, and TS Jessop (2013). “Monitoring the ungulate prey of the Komodo dragon *Varanus komodoensis*: distance sampling or faecal counts?” English. In: *Wildlife Biology* 19.2, pp. 126–137. URL: <http://dx.doi.org/10.2981/11-098>.
- Barry, Simon C. and A. H. Welsh (2001). “Distance Sampling Methodology”. English. In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 63.1, pp. 31–53. URL: <http://www.jstor.org/stable/2680632>.
- Buckland, S. T. (1987). “On the variable circular plot method of estimating animal density”. English. In: *Biometrics* 43.2, pp. 363–384. URL: <http://www.jstor.org/stable/2531819>.
- (2006). “Point-Transect Surveys for Songbirds: Robust Methodologies (Muestreos en Transectos Puntuales para Aves Canoras: Metodologías Robustas)”. English. In: *The Auk* 123.2, pp. 345–357. URL: <http://www.jstor.org/stable/4090665>.
- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas (2001). *Introduction to distance sampling: estimating abundance of biological populations*. English. Oxford; New York: Oxford University Press. ISBN: 019850649X; 9780198509271; 9780198506492; 0198509278.
- Burnham, Kenneth P., David R. Anderson, and Jeffrey L. Laake (1980). “Estimation of Density from Line Transect Sampling of Biological Populations”. English. In: *Wildlife Monographs* 72, pp. 3–202. URL: <http://www.jstor.org/stable/3830641>.
- Cox, Martin J., David L. Borchers, David A. Demer, George R. Cutter, and Andrew S. Brierley (2011). “Estimating the density of Antarctic krill (*Euphausia superba*) from multi-beam echosounder observations using distance sampling methods”. English. In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 60.2, pp. 301–316. URL: www.summon.com.
- Durant, Sarah M., Meggan E. Craft, Ray Hilborn, Sultana Bashir, Justin Hando, and Len Thomas (2011). “Long-term trends in carnivore abundance using distance sampling in Serengeti National Park, Tanzania”. English. In: *Journal of Applied Ecology* 48.6, pp. 1490–1500. URL: <http://dx.doi.org/10.1111/j.1365-2664.2011.02042.x>.
- Engbring, John, Fred L Ramsey, and Valerie J. Wildman (1986). “Micronesian forest bird survey, 1982: Saipan, Tinian, Agiguan, and Rota”. In: URL: <http://hdl.handle.net/2027/coo.31924062784776>.
- Isaac, Nick J. B., Katie L. Cruickshanks, Ann M. Weddle, J. Marcus Rowcliffe, Tom M. Brereton, Roger L. H. Dennis, David M. Shuker, and Chris D. Thomas (2011). “Distance sampling and the challenge of monitoring butterfly populations”. English. In: *Methods in Ecology and Evolution* 2.6, pp. 585–594. URL: <http://dx.doi.org/10.1111/j.2041-210X.2011.00109.x>.
- Laake, J. L., B. A. Collier, M. L. Morrison, and R. N. Wilkins (2011). “Point-Based Mark-Recapture Distance Sampling”. English. In: *Journal of Agricultural, Biological, and Environmental Statistics* 16.3, pp. 389–408. URL: <http://dx.doi.org/10.1007/s13253-011-0059-5>.

- Mack, Y. P. and Pham X. Quang (1998). “Kernel Methods in Line and Point Transect Sampling”. English. In: *Biometrics* 54.2, pp. 606–619. URL: <http://www.jstor.org/stable/3109767>.
- Marcus Rowcliffe, J., Chris Carbone, Patrick A. Jansen, Roland Kays, and Bart Kranstauber (2011). “Quantifying the sensitivity of camera traps: an adapted distance sampling approach”. English. In: *Methods in Ecology and Evolution* 2.5, pp. 464–476. URL: <http://dx.doi.org/10.1111/j.2041-210X.2011.00094.x>.
- Pederson, Åshlid Ø., Bård-Jørgen Bårdsen, Nigel G. Yoccoz, Nicolas Lecomte, and Eva Fuglei (2012). “Monitoring Svalbard Rock Ptarmigan: Distance Sampling and Occupancy Modeling”. English. In: *The Journal of Wildlife Management* 76.2, pp. 308–316. URL: <http://dx.doi.org/10.1002/jwmg.276>.
- Quang, Pham X. (1993). “Nonparametric Estimators for Variable Circular Plot Surveys”. English. In: *Biometrics* 49.3, pp. 837–852. URL: <http://www.jstor.org/stable/2532203>.
- Ramsey, Fred L and J Michael Scott (1979). “Estimating population densities from variable circular plot surveys”. In: *Sampling Biological Populations* 5.
- (1981). “Analysis of bird survey data using a modification of Emlen’s method”. In: *Studies in Avian Biology* 6, pp. 483–487.
- Reynolds, R. T., J. Michael Scott, and R. A. Nussbaum (1980). “A Variable Circular-Plot Method for Estimating Bird Numbers”. English. In: *The Condor* 82.3, pp. 309–313. URL: <http://www.jstor.org/stable/1367399>.
- Thompson, Steven K. (2012). *Sampling*. English. Vol. 755.; 755. Hoboken, N.J: John Wiley & Sons. ISBN: 1118162935; 9781118162965; 9781118162934; 111816296X. URL: www.summon.com.