Characterizing the "goodness" of a variety of sampling schemes *(Draft: April 7, 1999)*

Introduction

We've talked about a variety of sampling schemes but have not yet provided quantitative yardstick for characterizing these schemes. Below I construct a scenario that will allow us to focus on specific monitoring design issues and characterize the strengths and weaknesses of a variety of options.

Based on that scenario there are some definite conclusions that can be made, however:

1. If no along-shore variability in density is considered, then when using fixed-width transects there are closed-form solutions for the mean and variance of the estimated density, *i.e.*, no simulations are needed.

2. Under the condition of no along-shore variability, non-parallel-to-shore transects (perpendicular- and angular-to-shore transects) can have far less variability than parallel-to-shore transects.

3. Along-shore variability needs to be considered before we can make intelligent decisions about transect orientations. And more specifically (as Tim has mentioned before) spatial variability needs to be considered.

Definitions

First we define the target population and the parameters of interest:

- **Target Population.** This is an area off the coast of a recovery zone from d_1 to d_2 meters offshore over the time period of t_1 days to t_2 days past January 1 for a specified set of years. (Modifications of the definition will certainly need to be made for Recovery Plan Zone 1.)

- **Parameter(s) of Interest.** We want to estimate the mean density of marbled murrelets in the target population (averaged over space and time) for each of the specified set of years.

Next we define how the birds are distributed on the water over space and time in the target population for a particular year. As is standard in many initial simulations of counts of objects, we let the number of birds found in any area to follow a Poisson distribution. We allow for the Poisson parameter to vary with the distance from shore but assume that it is constant along the shore and constant throughout the target population time period of each year.

That offshore density function is labeled g and has a mean density given by

$$l = \int_{d_1}^{d_2} g(x)dx \div (d_2 - d_1)$$

Because we've assumed no variability in density in time or along the shore, then l is the parameter of interest described above. If we knew l, then we would be done. We wouldn't need to construct a sampling frame, fixed-width or line transects, a randomization scheme, or anything else. But, of course, we'll need to estimate l.

It will also be useful to talk about the variability in density and calculate the variance of g:

$$v = \int_{d_1}^{d_2} \left(g(x) - l\right)^2 dx \div (d_2 - d_1)$$

Now we consider 4 combinations of sampling frames and sampling schemes. We'll calculate the mean and variance of the estimate of density based on fixed-width strip transects. This will show us differences in biases and precision among the 4 combinations and hopefully suggest how we might proceed.

1. **Perpendicular**. A sample of n fixed-width transects of width w and length $l = d_2 - d_1$ perpendicular to the shore are chosen at random along the shore and at a random day in the population timeframe.

2. **Parallel**. A sample of n fixed-width transects of width w and length l parallel to the shore each with a random starting point along the shore, a random distance from shore, and on a random day in the population timeframe.

3. **Fixed distance *x*, parallel.** A sample of n fixed-width transects of width w and length l parallel to the shore each at x meters from shore with a random starting point along the shore, and on a random day.

4. **Angular.** A sample of n fixed-width transects oriented at a fixed angle q away from being perpendicular to shore of width w and length $l = (d_2 - d_1)/\sin q$ with the starting points chosen at random along the shore and on random days in the population timeframe.

Results

For each of these combinations the total area of transects is labeled $A = nwl$. The estimate of density will be the total number of birds observed divided by the total transect area covered. See below for a table containing the biases and variances of each of the estimators.

Transect Type	Mean	Bias	Variance
1. Perpendicular	l	0	l / A
2. Parallel	l	0	$l / A + v / n$
3. Fixed distance *x*, parallel	$g(x)$	$l - g(x)$	$g(x) / A$
4. Angular	l	0	l / A

All of the schemes that have each point in space and time with a positive probability of being selected result in zero bias. The Parallel scheme samples the variability at different distances only between samples and ends up with a larger variance associated with the variability of the offshore density.

If we include along-shore variability in our model, then the variability for the Perpendicular and Angular scheme will certainly increase. But (depending on the scale of the along-shore variability), we wouldn't expect the Angular variability to increase as much.

Appendix: Technical details

Suppose the number of birds detected in a specific area follows a Poisson distribution. We allow for the Poisson parameter to vary depending on the distance from shore.

Let the Poisson intensity function (*i.e.,* expected number of birds per unit area) be $g(x)$ where x is the distance from shore with $d_1 \le x \le d_2$) where the distances from shore d_1 and d_2 delineate the statistical population of interest).

Consider the following fixed-width transect orientations:

1. Transects perpendicular to shore of length $l = d_2 - d_1$ and width w. There will be n transects chosen randomly from along the coastline. The i-th transect will have the number of observed birds, Y_i, following a Poisson distribution with mean

$$
\begin{aligned}
m &= \text{transect area } \cdot \text{ mean density} \\
&= lw \cdot \int_{d_1}^{d_2} g(x)dx/(d_2 - d_1) \\
&= lw\,l
\end{aligned}
$$

   The sample mean count for the n transects will have a mean of m and a variance of m/n. Translating this into an estimate for the mean density of birds we have

$$\hat{l} = \text{sample mean count } \div \text{ transect area} = \sum_{i=1}^{n} Y_i / n \div (lw)$$

$$E(\hat{l}) = \frac{m}{lw} = \frac{lw\,l}{lw} = l$$

$$V(\hat{l}) = \frac{m}{n} \cdot \frac{1}{(lw)^2} = \frac{lw\,l}{n} \cdot \frac{1}{(lw)^2} = \frac{l}{lwn} = \frac{l}{wL} = \frac{l}{A}$$

   where $L = nl$ is the total transect length and $A = lwn$ is the total transect area.

2. Consider transects of length l and width w parallel to shore with random starting point along the shore with a random distance away from shore. The random distance away from shore is chosen from a uniform distribution between distances d_1 and d_2.

   The Poisson intensity parameter depends on the random distance from shore. We have the joint probability of observing y birds on a randomly selected transect with the transect distance from shore, X, being between a distance of x and $x + h$ is

$$
\begin{aligned}
\Pr(Y = y, x \le X \le x+h) &= \Pr(Y = y \mid x \le X \le x+h) \cdot \Pr(x \le X \le x+h) \\
&= \Pr(Y = y \mid x \le X \le x+h) \cdot \frac{h}{d_2 - d_1}
\end{aligned}
$$

Divide both sides by *h* and pass to the limit as $h \to 0$ gives us

$$
\lim_{h \to 0} \frac{1}{h} \Pr(Y = y, x \le X \le x+h) = \frac{1}{d_2 - d_1} e^{-lwg(x)} \frac{(lwg(x))^y}{y!}
$$

The unconditional probability of observing *y* birds is therefore

$$
\Pr(Y = y) = \int_{d_1}^{d_2} \frac{1}{d_2 - d_1} \cdot e^{-lwg(x)} \frac{(lwg(x))^y}{y!} dx
$$

The mean number of birds observed on such a randomly selected transect is

$$
\begin{aligned}
E(Y) &= \sum_{y=0}^{\infty} y \int_{d_1}^{d_2} \frac{1}{d_2 - d_1} \cdot e^{-lwg(x)} \frac{(lwg(x))^y}{y!} dx \\
&= \int_{d_1}^{d_2} \frac{1}{d_2 - d_1} \cdot \left[\sum_{y=0}^{\infty} y e^{-lwg(x)} \frac{(lwg(x))^y}{y!} \right] dx \\
&= \int_{d_1}^{d_2} \frac{1}{d_2 - d_1} \cdot lwg(x) dx \\
&= \frac{lw}{d_2 - d_1} \int_{d_1}^{d_2} g(x) dx \\
&= lwl = m
\end{aligned}
$$

and the variance is

$$
\begin{aligned}
V(Y) &= E(Y^2) - \left(E(Y)^2\right) \\
&= \sum_{y=0}^{\infty} y^2 \int_{d_1}^{d_2} \frac{1}{d_2 - d_1} \cdot e^{-lwg(x)} \frac{(lwg(x))^y}{y!} dx - m^2 \\
&= \int_{d_1}^{d_2} \frac{1}{d_2 - d_1} \cdot \left[\sum_{y=0}^{\infty} y^2 e^{-lwg(x)} \frac{(lwg(x))^y}{y!} \right] dx - m^2 \\
&= \int_{d_1}^{d_2} \frac{1}{d_2 - d_1} \cdot \left[lwg(x) + (lwg(x))^2 \right] dx - m^2 \\
&= lwl + (lw)^2 (v + l^2) - (lwl)^2 \\
&= lwl + l^2 w^2 v
\end{aligned}
$$

Translating this into an estimate for the mean density of birds we have

$$\hat{l} = \text{sample mean count } \div \text{ transect area} = \sum_{i=1}^{n} Y_i / n \div (lw)$$

$$E(\hat{l}) = lwl \cdot \frac{1}{lw} = l$$

$$V(\hat{l}) = \frac{lwl + l^2 w^2 v}{n} \cdot \frac{1}{(lw)^2} = \frac{l + lwv}{n} \cdot \frac{1}{lw} = \frac{l}{A} + \frac{v}{n}$$

So here we also have an unbiased estimate but the variance is larger by the variability in the Poisson intensity across the distances from shore. In addition, that additional variability is only reduced by taking more transects rather than taking more transect length or area.

When there's variability in the density perpendicular to the transect orientation we pay for that with an increased variance for a fixed total transect length. Transects perpendicular to shore will robust against variability in that same direction but not robust for variability along shore. Similar statements can be made about transects parallel to shore.

To guard against variability in both directions, it would seem advisable to choose transects that include both parallel to shore and perpendicular to shore orientations. Maybe we were too hasty in getting rid of zigzags?

My objection to zigzags were more based on the fact that they were completely arbitrarily chosen. If random selections of a standard sized zigzag would be made, then we would have a robust sampling scheme.

3. Consider transects of length l and width w parallel to shore with random starting point along the shore with a fixed distance x away from shore.

   All transects will have a mean density of $g(x)$. Both the mean and variance of the number of birds per transect will be $lwg(x)$. So the mean and variance of the estimate of density will be $g(x)$ and $g(x)/A$.

4. A sample of n fixed-width transects oriented at a fixed angle q away from being perpendicular to shore of width w and length $l = (d_2 - d_1)/\sin q$ with the starting points chosen at random along the shore and on random days in the population timeframe.

   The number of birds in a single transect will have a Poisson distribution with mean and variance lwl as such transects completely and identically cover the offshore gradient in density. The mean and variance of the estimated density will be l and l/A, respectively.