Estimating a Resource Selection Function With Line Transect Sampling

BRYAN F.J. MANLY   bmanly@west-inc.com
Western EcoSystems Technology Inc. 2003 Central Avenue, Cheyenne, Wyoming 82001 USA

Abstract. A resource selection probability function is a function that gives the probability that a resource unit (e.g., a plot of land) that is described by a set of habitat variables X_1 to X_p will be used by an animal or group of animals in a certain period of time. The estimation of a resource selection function is usually based on the comparison of a sample of resource units used by an animal with a sample of the resource units that were available for use, with both samples being assumed to be effectively randomly selected from the relevant populations. In this paper the possibility of using a modified sampling scheme is examined, with the used units obtained by line transect sampling. A logistic regression type of model is proposed, with estimation by conditional maximum likelihood. A simulation study indicates that the proposed method should be useful in practice.

Keywords: bootstrap resampling, distance sampling, logistic regression, resource selection by animals.

1.   Introduction

A resource selection probability function (RSPF) gives the probabilities of use for different types of resource units, where these units might, for example, be plots of land or items of food. Usually these functions are estimated by taking two random samples, one of which is of used units and the other either of unused units or of all available units. Differences between the two types of sample then reflect the nature of the RSPF, which can be estimated by logistic regression and other methods (Manly et al. [6]).

  In principle, other more complicated sampling designs may also be used to gather the data for RSPF estimation, but this topic has not received much attention to date. In this paper one such possibility is considered, in which line transect sampling is used to gather the data for the ”used” sample of resource units. This introduces the problem that the probability of detecting and recording a used unit is likely to depend on how far it is from the transect line, and possibly also on the type of habitat that the unit is in. Hence the usual methods for estimating a RSPF no longer apply.

The combination of having a line transect sample of used units and a random sample of available units does not seem to have been considered before, although methods for allowing for the probability of detection of a used unit to depend on the type of unit that it is have been used with ordinary line transect sampling (Beavers and Ramsey [1]). The advantage of using the two types of sampling that are discussed in the present paper is that it is possible to estimate both the RSPF and the total number of animals using the units.

In the next section of this paper a justification for using logistic regression for estimating a RSPF is given when the data available consists of a random sample of used units and a random sample of available units. Section 3 then suggests how this approach can be modified with data collected from line transect sampling, leading to a logistic regression type of model, with estimation of both the RSPF and the number of individuals in the area being studied. The properties of this model are examined by a simulation study in Section 4, and finally conclusions are summarised in Section 5.

2. Estimating a Resource Selection Probability Function From Random Samples

Suppose that there is a set of N resource units (e.g., plots of land) available for use by either a single animal or a group of animals, with each unit being described by the values that it possesses for some variables $X_1, X_2, ... X_p$ (e.g. the altitude, the percentage of the unit covered by grasslands, or the density of a certain type of tree). Suppose also that the probability of use for one of the resource units depends on the values that the unit possesses for the X variables. Then the RSPF gives this probability, and its value for the ith resource unit can be written as $w^*(x_i)$, where this unit has the values $x_i = (x_{i1}, x_{i2}, ..., x_{ip})$ for the X variables.

Next, assume that the sampling scheme is such that every available unit has a probability P_a of being sampled, and every used unit has a probability P_u of being sampled. Further, assume that the available units are sampled first, without replacement. In that case, the probability of a unit being recorded as used is $(1 - P_a)\, w^*(x_i)P_u$ and the probability of a unit being in either the available or the used sample is

$$\text{Prob}(i\text{th unit sampled}) = P_a + (1 - P_a)w^*(x_i)P_u. \tag{1}$$

It then follows that the probability that the ith unit is in the used sample, given that it is in one of the two samples, is

$$\text{Prob(ith unit used} \mid \text{sampled)} = \frac{\text{Prob(used and sampled)}}{\text{Prob(sampled)}}$$

$$= \frac{(1 - P_a)w^*(x_i)P_u}{P_a + (1 - P_a)w^*(x_i)P_u}. \tag{2}$$

It is convenient at this point to let the RSPF take the particular form

$$w^*(x_i) = exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip}), \tag{3}$$

where the argument of the exponential function must be negative. Letting $\theta(x_i)$ denote the probability on the left-hand side of equation (2), it then follows that

$$\theta(x_i) = \frac{exp\{log_e[P_u(1 - P_a)/P_a] + \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip}\}}{1 + exp\{log_e[P_u(1 - P_a)/P_a] + \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip}\}}. \tag{4}$$

This is a logistic regression equation in which the parameter β_0 is modified by the correction term $log_e[P_u(1 - P_a)/P_a]$ to allow for available and used resource units being sampled with different probabilities.

Assuming independence of observations, the probability of observing resource unit i as used is $\theta(x_i)$ and the probability of observing it as available is $1 - \theta(x_i)$. Let y_i be an indicator of whether a sampled unit was used. That is, $y_i = 0$ if sampled unit i came from the available sample, and $y_i = 1$ if unit i came from the sample of used units. Using the y values as responses, any standard logistic regression programs can be used to estimate the coefficients $\beta_0, \beta_1, ..., \beta_p$ and their approximate variances.

The fact that the constant in the logistic regression is $log_e[P_u(1-P_a)/P_a] + \beta_0$ means that if the P_u and P_a are known then the parameter β_0 in the RSPF can be estimated by subtracting $log_e[P_u(1 - P_a)/P_a]$ from the estimated constant in the logistic regression equation. If P_u and P_a are not known then β_0 cannot be estimated, but it is still possible to estimate the resource selection function (RSF)

$$w(x) = exp\{\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p\} \tag{5}$$

and use this to compare resource units. In practice there is little loss from using the RSF rather than the RSPF because the RSF provides values that are proportional to probabilities of use. This is all that is needed to determine which types of resource unit are apparently favoured by the animals being studied, produce maps of habitat use, etc. (Manly et al. [6]).

3.   Estimation with Distance Sampling

It will now be assumed that the situation described in the previous section is modified because the used sample is obtained by line transect sampling, although the available sample is randomly selected, as before. It is envisaged that the available sample might be collected by recording habitat information on randomly selected resource units during the line transect sampling, or possibly at a different time. The available sample might even be obtained from the data in a geographical information system.

The probability of resource unit i being used is still given by the RSPF, $w^*(x_i)$, which is assumed to be unrelated to the distance of the resource unit from the transect line. However, the probability of the use being recorded will generally depend on the distance, d_i, of the unit from the line, and may also depend on the nature of the unit as defined by the variables in the RSPF. Hence, this probability of recording unit i if it is used will be denoted by $g(d_i, x_i)$. The probability of the ith unit being either included in the available sample or in the used sample then becomes

$$P_a + (1 - P_a)w^*(x_i)g(d_i, x_i), \tag{6}$$

assuming, as before, that the available units are sampled first with probability P_a, without replacement. Consequently, the probability of the ith unit appearing in the used sample, given that it is in one of the samples, is

$$\theta(x_i, d_i) = (1 - P_a)w^*(x_i)g(d_i, x_i)/\{P_a + (1 - P_a)w^*(x_i)g(d_i, x_i)\} \tag{7}$$

At this stage some specific assumptions about the form of the functions become useful. It will therefore be assumed that the RSPF is as given in equation (3), while the other function takes the form

$$g(d_i, x_i) = exp\{-d_i(\gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + ... + \gamma_p x_{ip})\}. \tag{8}$$

These assumed parametric forms for the functions are arbitrary, and are used here because they are convenient to develop the ideas being presented. Other functions could well be used instead, should this be necessary.

What can be noted is that equation (3) for the RSPF has often been used in the past, and the important characteristic is that it is always positive. Equation (8) for the detection function is also reasonable because $g(0, x_i) = 1$, so that all used units on the transect line are assumed to be recorded, and as the distance from the line increases the probability of recording a use decreases for all types of units.

Substituting into equation (7) using equations (3) and (8) leads to

$$\theta(x_i, d_i) = \frac{A_i}{1 + A_i} \tag{9}$$

which is essentially a non-linear logistic regression model, with $2p + 1$ unknown parameters, which are $\beta_0, \beta_1, ..., \beta_p, \gamma_0, \gamma_1, ...$ and γ_p, where A_i denotes

$$exp\{\beta_0 + log_e[\frac{(1 - P_a)}{P_a}] - d_i\gamma_0 + (\beta_1 - d_i\gamma_1)x_{i1} + ... + (\beta_p - d_i\gamma_p)x_{ip}\}.$$

If the value of P_a is known then this can be allowed for when fitting equation (9) to data. If it is not known then β_0 and $log_e[(1-P_a)/P_a]$ are confounded and only the combined parameter $\beta_0' = \beta_0 + log_e[(1 - P_a)/P_a]$ can be estimated.

Suppose that the units in the available sample are numbered from 1 to n_a, and the units in the used sample are numbered from n_a+1 to n_a+n_u. Then the likelihood function for the full sample is the product of the probabilities of the units being in the sample that they were observed in, given that they were in one of the two samples. This likelihood function is therefore

$$L = \Pi_{i=1}^{n_a}\{1 - \theta(x_i, d_i)\}\Pi_{i=1}^{n_a+n_u}\theta(x_i, d_i). \tag{10}$$

This can be maximized numerically to obtain the maximum likelihood estimators of the unknown parameters. For the results presented later this was done using the algorithm AMOEBA of Press et al. [7].

Transect sampling is usually carried out in order to estimate the number of individuals in the sampled area. In the present context this means estimating the total number of used units in the area, assuming that each used resource unit contains only one animal. There are two obvious moment estimators that might be used for this purpose when P_a is known. First, if the average value of the estimated RSPF for the n_a units in the sample of available units is $\hat{w}_a$, then

$$\hat{E}_a = N\hat{w}_a \tag{11}$$

is an estimator of the total number of used units. Second, if the ith of the n_u units recorded as being used is at a distance of d_j from the transect line, then a Horvitz-Thompson estimator of the total number of used units is

$$\hat{E}_u = \Sigma_{j=1}^{n_u}[1/\hat{g}_j] \tag{12}$$

(Horvitz and Thompson [4]; Thompson [8], p. 49), where $\hat{g}_j$ is the estimated probability of use being detected for the jth unit.

A simple way to assess the accuracy of the estimators of the parameters of the RSPF, the distance function, and the two estimators of population size, is by bootstrap resampling. Use a bootstrap set of data obtained by resampling the observed available units to obtain a new sample of available

units with the same size as the observed sample, and independently resample the used units to get a new sample of used units, again with the same size as the observed sample (Buckland et al. [2], p. 94). There are other ways of doing the bootstrap resampling, but only this simple approach is considered here. Each bootstrap set of data will provide new estimates, and the variances of these estimates are approximations to the variances that would be obtained by repeating the real world sampling process.

An obvious question of interest is how to combine the two estimators of population size in order to obtain a single improved estimator. Again adopting a simple approach suggests using the linear combination with the minimum variance, which is

$$\hat{E} = \frac{\{Var_B(\hat{E}_u) - Cov_B(\hat{E}_a, \hat{E}_u)\}\hat{E}_a + \{Var_B(\hat{E}_a) - Cov_B(\hat{E}_a, \hat{E}_u)\}\hat{E}_u}{Var_B(\hat{E}_a) + Var_B(\hat{E}_u) - 2Cov_B(\hat{E}_a, \hat{E}_u)}$$

where $Var_B(\hat{E}_a)$, $Var_B(\hat{E}_u)$ and $Cov_B(\hat{E}_a, \hat{E}_u)$ denote variances and a covariance estimated by bootstrapping (Manly [5], Appendix A6). However, it turns out that because of the high correlation between the estimators $\hat{E}_a$ and $\hat{E}_u$ this often results in one of these estimators having a negative weighting. Therefore it seems better not to take the covariance into account and use

$$\hat{E} = \frac{Var_B(\hat{E}_u)\hat{E}_a + Var_B(\hat{E}_a)\hat{E}_u}{Var_B(\hat{E}_a) + Var_B(\hat{E}_u)} \tag{13}$$

instead. The variance is then estimated by

$$\begin{aligned} Var(\hat{E}) &= A^2 Var_B(\hat{E}_a) + (1-A)^2 Var_B(\hat{E}_u) \\ &\quad - 2A(1-A)Cov_B(\hat{E}_a, \hat{E}_u), \end{aligned} \tag{14}$$

where $A = Var_B(\hat{E}_u)/\{Var_B(\hat{E}_a) + Var_B(\hat{E}_u)\}$.

Although these equations have been described for estimates of population size, they can also be used with estimates of the logarithm of population size, as long as the bootstrap variances and covariance are estimated for these transformed values.

4.   Simulation Studies

Some simulation experiments have been carried out to investigate the use of the method described in the previous section. Unfortunately, these experiments have had to be based on an artificial situation because no real data are available where a sample of available units has been taken at the same times as a sample of used units gathered using distance sampling.

With the artificial situation the transect line is 5,000 units long, with 5 units on the left and 5 units on the right of the line, as shown in Figure 1. Moving away from the transect line the centre of the units are at distances of 0.1, 0.3, 0.5, 0.7 and 0.9 units from the line.
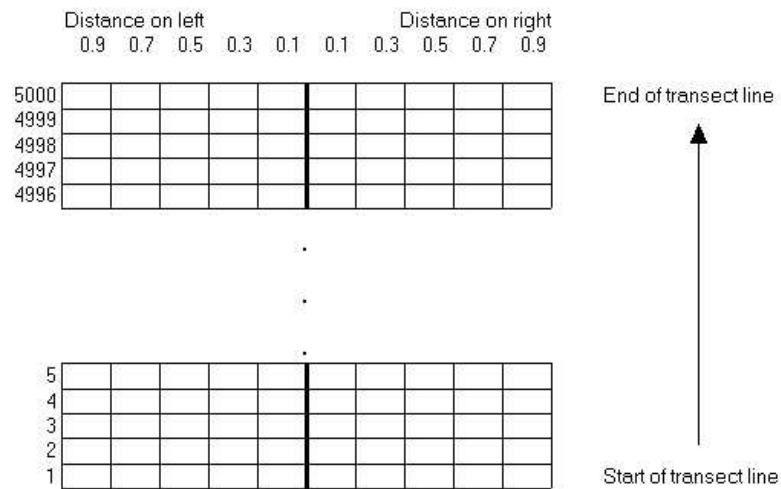

Figure 1. The artificial situation considered where there are 50,000 resource units in a 10 by 5,000 array.

Initially, the RSPF for this population was made equal to a function of a single variable X_1, with the form

$$w^*(x_1) = exp(\beta_0 + 0.3x_1). \tag{15}$$

Changing the value of β_0 has the effect of changing the probability of use, and this parameter was adjusted to obtain various expected sample sizes of used units for the simulation experiment.

The values for X_1 were derived from 5,000 sequential values for the octals of sea ice observed from aerial surveys of Pacific walrus conducted over many years (Joel Garlich-Miller, personal communication). These 5,000 values were associated with the 5,000 positions for resource units along the transect line (Figure 1) with the value of X_1 for a resource unit being the associated octal value plus a random value from an exponential distribution

with a mean of one. This resulted in values for X_1 with the type of spatial correlation that is likely to occur with measurements on resource units.

Initially the function for the probability of detection of a used unit at distance d from the transect line was set at

$$g(d, x_1) = exp\{-d(1 + \gamma_1 x_1)\}$$

For most simulations the parameter γ_1 was set equal to zero, in which case the probability of detecting a used unit varied from $exp(-0.1) = 0.90$ for the closest units to the transect line to $exp(-0.9) = 0.41$ for the units furthest from the transect line.

To obtain a simulated set of data, each of the 50,000 resource units in the population first had the opportunity to be in the available sample, with probability P_a. If the unit was not selected for the available sample, then it was recorded as used with probability $w^*(x_1)g(d, x_1)$.

4.1.   Main Simulation Study A

The main simulation study was designed to determine the sample size requirements for estimation when there is only a single X variable available to describe resource units. To this end, nine scenarios were simulated, with a two factor design. The factors and their levels were as follows:

- Factor A: The expected number of available units sampled, with the levels: 250 units, 1000 units, and 4000 units.

- Factor B: The expected number of used resource units sampled, with the levels: 100, 500 and 1000 units.

The levels for factor A were obtained by setting the sampling probability P_a to the appropriate level. Similarly, the levels for factor B were obtained by adjusting the parameter β_0 in equation (15) by a process of trial and error, in order to obtain the required values. One hundred sets of data were generated for each factor combination, with 100 bootstrap resamples used to estimate variances.

For estimation, the model assumed for the RSPF was

$$w^*(x_1) = exp(\beta_0 + \beta_1 x_1).$$

This is the correct form and the effectiveness of the estimation is indicated by the distributions of the estimates of β_0 and β_1 being unbiased, with small variances.

The model assumed for the detection function was

$$g(d, x_1) = exp\{-d(\gamma_0 + \gamma_1 x_1)\}.$$

Again, this is the correct form and successful estimation requires that the distributions of the estimates of γ_0 and γ_1 are unbiased with small variances.

Finally, estimates of population size were be calculated using equations (11) and (12), and combined using equation (13). For the simulations, logarithms of these population size estimates were considered rather than the population size estimates themselves. This is because the logarithms of estimates were found on inspection to have distributions that were closer to normal than the distributions of the population size estimates themselves. As noted in Section 3, the logarithms of size can be combined using equation (13) providing that all references to sizes are replaced by references to the logarithms of sizes.

As far as estimates of the logarithm of size are concerned, the main concern is that, like the estimates of other parameters, they are unbiased with acceptably small variances.

Figure 2 summarises the results obtained for the estimation of the parameters β_0 and β_1 for the RSPF. For β_0 there is little evidence of bias, and it is apparent that increasing the number of available units has little effect on the variation of estimates when only 100 used units are expected, but has a substantial effect when more used units are expected. The same is also true with β_1, but in this case increasing the expected number of used units from 500 to 1,000 has also not had much effect on accuracy. It appears, therefore, that increasing either the available or used sample size considerably may not in itself necessarily give much reduction in the variation of estimates.

Figure 3 is similar to Figure 2, but is for the estimation of the parameters γ_0 and γ_1 of the detection function. Again it is noticeable that when the expected number of used units is only 100 the standard deviation of the estimates of γ_0 is not reduced much, if at all, by increasing the number of used units. Also, increasing the expected number of used units from 500 to 1,000 has had little effect for estimates of this parameter. These effects are not so clear for the estimates of γ_1, but still seem to be present.

The results for the estimation of the logarithm of the number of used units are summarised in Figure 4. In fact, E_a and E_u gave almost the same results for most sets of data, with E_u tending to give a consistently lower variance that E_a, but also having more evidence of bias. It was also found that combining the logarithms of estimates using equation (13) gave little improvement over using E_a directly. Consequently, it is not obvious
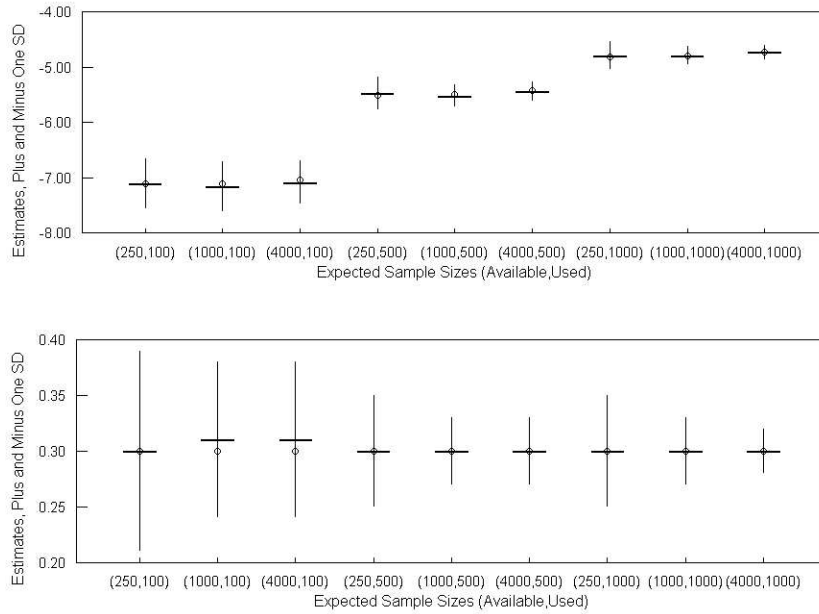
Figure 2. Summary of the results obtained for estimates of the parameters β_0 and β_1 of the resource selection probability function. For each combination of the expected number of available and used sample sizes, the horizontal line is the mean of the 100 estimates obtained, and the vertical line extends from the mean minus one standard deviation to the mean plus one standard deviation. The open circle is the true value for the parameter being estimated.

which estimator is the best to use. In terms of the mean square error the combined estimate seems to possibly be slightly better than its components.

For all estimators increasing the expected number of used units from 100 to 500 has reduced the standard deviation considerably. However, increasing the expected number of used units from 500 to 1,000 has not shown much more reduction, if any, in the standard deviations.

Some comments on the bootstrap estimation of variances are in order before leaving the consideration of the results of simulation study A. In brief, the bootstrap method worked reasonably well in the situation being considered, with the average values of bootstrap standard deviations for the estimators being close to the observed standard deviations from the independently generated sets of data.
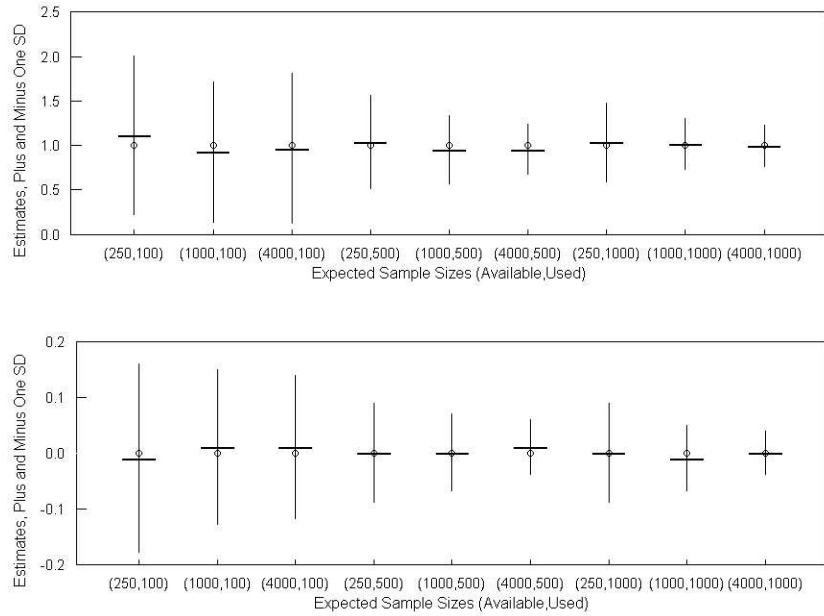
Figure 3. Summary of the results obtained for estimates of the parameters γ_0 and γ_1 of the detection probability function. For each combination of the expected number of available and used sample sizes, the horizontal line is the mean of the 100 estimates obtained, and the vertical line extends from the mean minus one standard deviation to the mean plus one standard deviation. The open circle is the true value for the parameter being estimated.

As an example, consider the estimation of β_0 for the situation where the expected number of available units sampled is 1,000 and the expected number of used units is 500. The standard deviation observed for the 100 estimates of this parameter was 0.20. The bootstrap estimates of the standard deviation from the same 100 sets of data had a mean of 0.21, with a range of from 0.15 to 0.26 and a standard deviation of 0.02.

4.2.   Further Simulation Studies

The situation considered in the main simulation study was the simplest possible, with just a single X variable. It is also interesting to see how the results change if there is a second variable of interest to be included
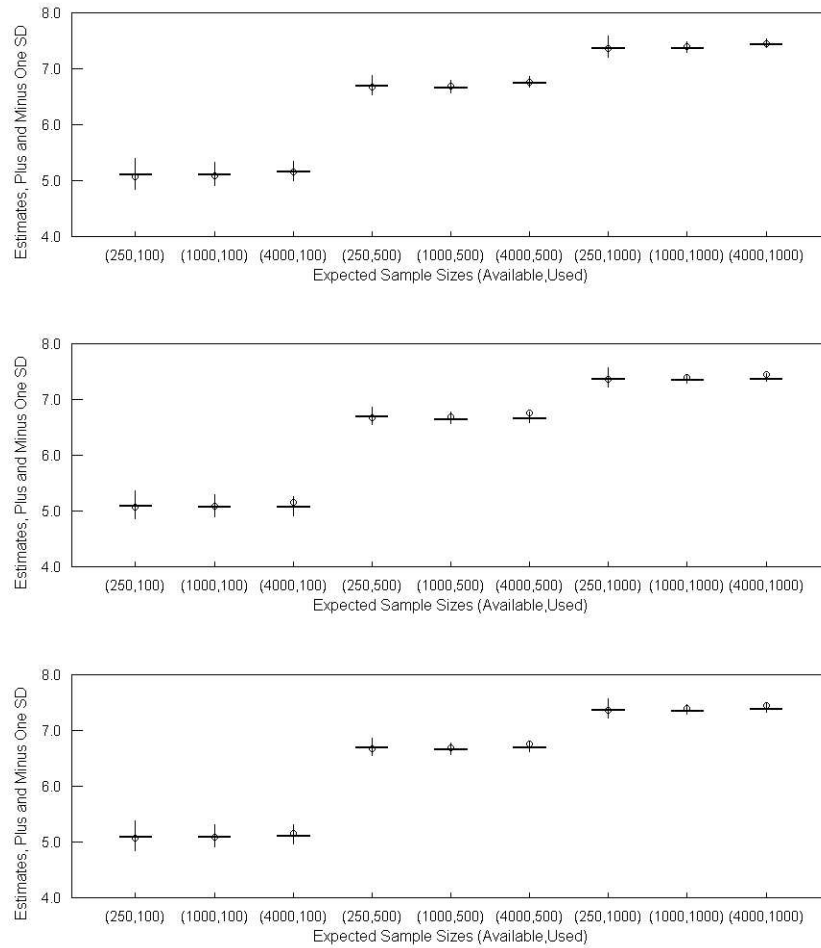
Figure 4. Summary of the results obtained for estimates of the logarithm of the total number of used units in the study area. For each combination of the expected number of available and used sample sizes, the horizontal line is the mean of the 100 estimates obtained, and the vertical line extends from the mean minus one standard deviation to the mean plus one standard deviation. The open circle is the true value for the parameter being estimated.

in the estimated RSPF and the estimated detection function, although in fact this second variable is not really part of these functions. This has

Table 1. Comparison of the means and standard deviations of estimates of parameters obtained under different conditions when the expected sample size for available units is 1,000 and for used units is 500. The variable X_2 had no effects in these simulations.

Parameter	True Value	Without estimating X_2 parameters		When estimating X_2 parameters	
		Mean of Est	Std. Dev.	Mean of Est	Std. Dev.
β_0	-5.49	-5.52	0.20	-5.53	0.25
β_1	0.30	0.30	0.03	0.30	0.04
β_2	0.00	-	-	0.00	0.04
γ_0	1.00	0.95	0.39	0.97	0.44
γ_1	0.00	0.00	0.07	0.00	0.07
γ_2	0.00	-	-	0.00	0.07
Estimates of the Logarithm of the Total Number of Used Units					
$log_e(E_a)$	6.69	6.68	0.12	6.69	0.11
$log_e(E_u)$	6.69	6.66	0.11	6.67	0.11
$log_e(E_c)$	6.69	6.67	0.11	6.68	0.11

been examined very briefly for one particular scenario, where the expected number of available units sampled is 1,000 and the expected number of used units sampled is 500.

The values of this second variable were the same as those for X_1, but in the reverse order down the transect. This resulted in there being a small negative correlation between X_1 and X_2 ($r = -0.13$) in the population of 50,000 resource units. However, for simulating data the coefficient of X_2 was zero in the RSPF and also zero in the detection function. Thus the second variable was actually redundant. This meant that the data already simulated with the required sample sizes could be used again, but now with coefficients for X_1 and X_2 estimated in the RSPF.

The outcome of estimating coefficients for the redundant variable was that the estimates of β_0, β_1, γ_0 and γ_1 became slightly more variable, but otherwise the distribution of estimates was very reasonable (Table 1, columns 5 and 6). Interestingly, the estimation of the number of used units was just as good as it was when X_2 was not included in the equation.

Finally, data were simulated with the RSPF and distance function depending on both X_1 and X_2. The coefficients for X_1 and X_2 were given opposite signs in the two functions with the idea that the preferred habitat might be in places where the animals are difficult to see. As for the other situations discussed in this section, the parameters were set so that

Table 2. Comparison of the means and standard deviations of estimates of parameters obtained when the expected sample size for available units is 1,000 and for used units is 500.

Parameter Parameter	True Value	Mean of Estimates	Standard Deviation For 100 Estimates	Mean of Boostrap Values
β_0	-4.63	-4.63	0.27	0.26
β_1	0.30	0.30	0.05	0.04
β_2	-0.30	-0.31	0.06	0.06
γ_0	2.00	2.01	0.59	0.57
γ_1	-0.10	-0.11	0.09	0.09
γ_2	0.10	0.11	0.12	0.13
Estimates of the Logarithm of the Total Number of Used Units				
$log_e(E_a)$	6.91	6.90	0.13	0.12
$log_e(E_u)$	6.91	6.88	0.13	0.12
$log_e(E_c)$	6.91	6.89	0.13	0.12

the expected sample size for available units was 1,000 and the expected sample size for used units, with 100 sets of data generated and analysed. This situation was considered to ensure that estimation is effective when the distance function depends on the X variables.

Table 2 summarises the results obtained. Basically, estimation of all the unknown parameters is essentially unbiased, with the size of the standard deviation of estimates being well estimated by bootstrapping, at least on average. As before, there is little to choose between alternative estimators of the total number of used units.

5. Discussion

This has been a very preliminary examination of the extension of logistic regression methods for estimating RSFs to accommodate distance sampling of used units. The primary motivation was a desire to see whether this is a potentially useful method for combining the estimation of the size of a population with a study of resource selection.

It is clear from the limited simulation results that have been presented that the proposed model may be estimated reasonably well with moderate sized samples. Therefore, it may be that the general approach that has

been proposed does have some practical value, particularly if it was used with an appropriate model selection procedure such as the use of some version of Akaike's information criterion (Burnham and Anderson [3]).

Apart from model selection procedures, there are several aspects of the model that should be examined further in the future, including:

- alternative parameterizations of the resource selection function and the detection function;

- the comparison of the fit of alternative models using likelihood ratio methods;

- the relative amounts of effort that should go into the collection of the samples of used and available resource units;

- how to allow for situations where resource units are used by groups of individuals rather than single individuals;

- the need to ensure that values from the RSPF and detection function do not exceed one.

The last point is not likely to be a problem with most real sets of data because the probabilities of use are usually extremely small and the detection probability clearly decreases as resource units become further from the transect line. However, if necessary the model can just be changed so that probabilities greater than one cannot occur. For example, the RSPF can be assumed to take the form

$$w^*(x_i) = exp\{-exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip})\},$$

and the detection function the form

$$g(d_i, x_i) = exp\{-d_i exp(\gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + ... + \gamma_p x_{ip})\}.$$

In principle, using these alternative functions seems to introduce no new problems.

Finally, there is the question of what happens if a sample of used units is selected by line transect sampling without any allowance for visibility bias. It seems that this may not matter much providing that the detection function is the same for units with all types of habitat and transect lines are randomly placed because then the obtained sample will still be a random sample of used units. However, without actually estimating a detection function related to habitat variables there seems no way of being sure whether this is the case or not. Furthermore, for most animals it seems likely that the visibility will in fact vary with the habitat. If this is the case, and it is ignored, then the RSPF will be confounded with the detection function.

Acknowledgments

References

1. S. C. Beavers and F. L. Ramsey. Detectability analysis in transect surveys. *Journal of Wildlife Management* 62:948–57, 1998.

2. S. T. Buckland, D. R. Anderson, K. P. Burnham and J. L. Laake. *Distance Sampling: Estimating Abundance of Biological Populations.* London: Chapman and Hall, 1993.

3. K. P. Burnham and D. R. Anderson. *Model Selection and Inference.* New York: Springer-Verlag, 1998.

4. D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47:663–85, 1952.

5. B. F. J. Manly. *The Statistics of Natural Selection.* London: Chapman and Hall, 1985.

6. B. F. J. Manly, L. L. McDonald and D. L. Thomas. *Resource Selection by Animals: Statistical Design and Analysis for Field Studies.* London: Chapman and Hall, 1993.

7. W. H. Press, S. A. Teukolsky, W. T. Vetterling and D. F. Flannery. *Numerical Recipes in FORTRAN: the Art of Scientific Computing.* Cambridge: Cambridge University Press, 1992.

8. S. K. Thompson. *Sampling.* New York: John Wiley and Sons, 1998.