Double-Observer Line Transect Methods: Levels of Independence

Stephen T Buckland[1], Jeffrey L Laake[2] and David L Borchers[1]
[1]CREEM, University of St Andrews, Fife KY16 9LZ, Scotland
[2]NMML, Alaska Fisheries Science Center, NMFS, Seattle, WA 98115, USA

In line transect sampling, an observer travels along each of a number of lines, laid out according to some randomised (usually systematic random) scheme, and records each detected animal, together with its perpendicular distance from the line. One of the key assumptions of the method is that animals on the line are certain to be detected.

A number of authors have considered so-called double-observer or double-platform methods to extend line transect sampling to the case that not all animals on the line are detected. The double-observer data can be regarded as two-sample mark-recapture. However, heterogeneity in detection probabilities generates bias in abundance estimates, just as heterogeneity in capture probabilities generates bias in mark-recapture estimates of abundance.

In the absence of any heterogeneity in detection probabilities, we might assume that observer i detects any given animal in the surveyed strip with probability p_i, i = 1,2, and that the probability that both observers detect a given animal is $p_{12} = p_1p_2$. This is the 'full independence' assumption. However, in line transect sampling, we allow detection probability to fall off with distance y from the line so that $p_i = p_i(y)$. Thus it is natural to apply the full independence assumption at each distance from the line, so that for an animal at y, we assume $p_{12}(y) = p_1(y)p_2(y)$.

Laake (1999) argued that dependence was weaker close to the line, where probability of detection is greater, and proposed the 'point independence' assumption that $p_{12}(0) = p_1(0)p_2(0)$. Under this assumption, we can model both the unconditional detection function $p_i(y)$ and the conditional function $p_{i|i'}(y)$, i = 1,2, i' = 3 – i, which differ except at y = 0, and hence to estimate abundance without having to make the full independence assumption.

Although we can anticipate less dependence between detections on the line than at greater distances, unless detection on the line is certain, it seems possible that some dependence remains. We can replace the point independence assumption by the weaker assumption that, as detection probability tends to unity, dependence tends to zero (i.e. independence). We term this 'limiting independence'.

To implement limiting independence, we need a model for $\delta(y)$, where $p_{12}(y) = \delta(y)p_1(y)p_2(y)$ is the probability of detection by both observers at distance y. (Under full independence, $\delta(y)$ = 1 for all y, and under point independence, $\delta(0)$ = 1.) The constraints that $\delta(y)$ must tend to one as either $p_1(y)$ or $p_2(y)$ tend to one, and that all probabilities must lie in [0,1], lead to the following constraints on our model:

1. $\delta(y) \le U(y)$ where $U(y) = \min[1/p_1(y), 1/p_2(y)]$ (This ensures $p_{i|i'}(y) \le 1$)
2. $\delta(y) \ge L(y)$ where $L(y) = \max[0, \{p_1(y) + p_2(y) - 1\}/\{p_1(y).p_2(y)\}]$ (This ensures $p.(y) \ge 0$ where $p.(y) = p_1(y) + p_2(y) - p_{12}(y) = p_1(y) + p_2(y) - \delta(y).p_1(y).p_2(y)$ is the probability that at least one observer detects an animal at distance y.

Consider a logistic model of the form $$\log_e\left(\frac{\delta_0(y)}{1-\delta_0(y)}\right) = \alpha + \beta y + \log_e\left(\frac{1-L(y)}{U(y)-1}\right)$$

where $$\delta_0(y) = \{\delta(y) - L(y)\}/\{U(y) - L(y)\}$$

Then $\alpha = \beta = 0$ corresponds to full independence and $\alpha = 0$ to point independence. Hence this provides a framework for deciding the appropriate level of independence to assume.

To fit the model, we use a likelihood with two components, one corresponding to distances from the line of animals seen by at least one observer, and the other corresponding to the mark-recapture capture histories, where (1,1) corresponds to detection by both observers, (1,0) to detection by observer 1 only, and (0,1) by observer 2 only. Abundance is estimated using a Horvitz-Thompson-like estimator in which the inclusion probabilities are estimated from our model (Laake and Borchers, 2004).

Simulation study

We show in Table 1 results from analysing simulated data under limiting independence models with varying degrees of dependence. These results show that bias can be substantial for both full and point independence models. Strong correlation can occur between parameter estimates for limiting independence models; after rejecting the worst cases, some upward bias in the limiting independence estimates remains.

Table 1. Mean and standard error of 100 abundance estimates under full independence (FI), point independence (PI) and limiting independence (LI) models for eight simulation scenarios. Also given are model-averaged (MA) estimates, obtained by taking a weighted average of estimates from the above three models, using AIC weights. Where an LI model was deemed to be parameter-redundant, the weighted average was over the FI and PI models only. The mean and standard error for LI models is across only those runs for which the model was not deemed to be parameter-redundant. The number of runs (Nr) out of 100 contributing to the LI results under each scenario is shown. True abundance is 1000. *Bias significant at 5% level.

	FI		PI		LI			MA	
Scenario	Mean	s.e.	Mean	s.e.	Mean	s.e.	Nr	Mean	s.e.
1	607*	3	819*	6	1074*	14	95	1020	16
2	450*	1	692*	4	1037*	18	97	1020	18
3	499*	4	616*	6	971	21	60	685*	7
4	290*	3	454*	4	963	21	63	664*	21
5	512*	4	650*	6	1437*	124	76	1021	96
6	296*	3	465*	5	1149*	42	86	962	41
7	770*	2	915*	3	1064*	15	82	1001	14
8	689*	2	850*	4	1057*	11	94	1032*	12

Wooden stake data

Laake (1999) used data on a population of 150 wooden stakes to illustrate independence issues. Eight observers each surveyed the stakes. We consider analyses of data for each pair of observers. Observer was included as a factor in the logistic model. Results appear in Table 2.

Table 2. Estimates of abundance under full independence, FI ($\alpha = \beta = 0$), point independence, PI ($\alpha = 0, \beta$ unconstrained) and limiting independence, LI with $\alpha \ge 0, \beta \ge 0$. Covariate model 1 has covariate structure `observer + distance`, model 2 has structure `observer * distance`, and model 3 has structure `observer * (distance + distance²)`. 'Best' corresponds to the model with the smallest AIC (indicated by #), and 'MA' is the model-averaged estimate, using AIC weights. True abundance is 150.

		Cov model 1			Cov model 2			Cov model 3				
Obs1	Obs2	FI	PI	LI	FI	PI	LI	FI	PI	LI	Best	MA
1	2	136	139	156	132#	134	136	133	138	137	132	134
1	3	125	146	160	125	154	195#	120	145	156	195	183
1	4	127	158	215#	126	149	176	121	137	152	215	173
1	5	127	141#	154	126	136	141	128	140	144	141	137
1	6	126#	134	159	126	134	147	122	135	144	126	133
1	7	120	135	204	118	124	138	122#	125	127	122	126
1	8	136#	142	156	136	142	155	136	141	160	136	140
2	3	147	154	254	141#	150	157	143	154	171	141	147
2	4	86	131#	244	86	130	163	84	131	142	131	149
2	5	78	117	124	77	132#	135	75	143	148	132	134
2	6	105	149	239	105	142#	146	105	153	181	142	162
2	7	92	130#	252	92	129	137	88	127	138	130	139
2	8	107	144	197	105	129	131	105	126#	129	126	130
3	4	129	157	278	124	152	175#	124	148	162	175	163
3	5	143	164	246	141	156#	145	140	155	209	156	173
3	6	122	139	177	122	144	166#	122	150	156	166	154
3	7	130	166	294	127	157	213#	123	143	169	213	192
3	8	136	147	151	137	153#	172	135	153	159	153	153
4	5	105	156	182	105	157#	166	101	156	190	157	161
4	6	91	140	351#	93	130	178	94	137	136	351	215
4	7	99	137#	148	99	137	160	97	133	145	137	140
4	8	107	141	222	106	124	145#	109	124	130	145	148
5	6	115	167#	169	115	176	179	114	177	186	167	171
5	7	91	142#	147	91	145	150	87	143	146	142	144
5	8	117	148#	163	117	142	156	115	141	167	148	151
6	7	96	132	214#	97	126	130	99	120	132	214	164
6	8	106	134	159	106	132#	134	106	133	130	132	137
7	8	86	152	154	97	119	96	101	108#	111	108	118
Average		114	144	199	113	141	154	113	140	152	158	153
Std dev.		19	12	54	18	13	23	18	14	22	47	22

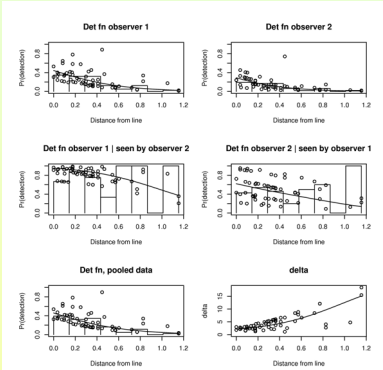

Fig. 1. Estimated detection functions for narwhals. Top: estimated unconditional detection functions for observer 1 (left) and 2 (right). Middle: conditional detection functions for observer 1 (left) and 2 (right). Bottom left: estimated detection function for the two observers combined. Bottom right: estimated $\delta(y)$, showing very strong dependence. The plotted curves correspond to the mean of recorded Beaufort values and cluster sizes, and the points show predicted values corresponding to each individual detection, conditional on the observed values of Beaufort and cluster size for that detection.

Discussion

The limiting independence formulation provides a framework for testing at what level independence can reasonably be assumed. It also allows estimation to take place under a very weak independence assumption – that independence occurs only in the limit as detection probability tends to unity. The simulation results show that such estimation is not always reliable. However, the limiting independence estimates of population size for the wooden stakes were on average closest to the true population size for two of the three models, and model-averaged estimates were remarkably close on average to the true abundance of 150 (Table 2). Methods are readily extended to allow the dependence to be a function of covariates other than distance, as is shown in Fig. 1, where probability of detection of narwhals was modelled as a function of distance, cluster size (i.e. number of animals in the group) and Beaufort sea state. See Buckland *et al*. (submitted) for more details.

References

Buckland, S.T., Laake, J.L. and Borchers, D.L. Double-observer line transect methods: levels of independence. *Submitted to Biometrics*.

Laake, J.L. (1999). Distance sampling with independent observers: reducing bias from heterogeneity by weakening the conditional independence assumption. Pp. 137-48 in *Marine Mammal Survey and Assessment Methods*, eds G.W. Garner, S.C. Amstrup, J.L. Laake, B.F.J. Manly, L.L. McDonald and D.G. Robertson. Balkema, Rotterdam.

Laake, J.L. and Borchers, D.L. (2004). Methods for incomplete detection at distance zero. Pp. 108-189 in *Advanced Distance Sampling*, eds S.T. Buckland, D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers and L. Thomas. Oxford University Press, Oxford.