COMBINING POPULATION DENSITY ESTIMATES IN LINE TRANSECT SAMPLING USING THE KERNEL METHOD

Patrick D. Gerard
Mississippi State University

and

William R. Schucany
Southern Methodist University

Combining Population Density Estimates in Line Transect Sampling Using the Kernel Method

Patrick D. Gerard

Experimental Statistics Unit, Box 9653, Mississippi State University, MS 39762

email : pdg1@ra.msstate.edu

William R. Schucany

Department of Statistical Science, Southern Methodist University, Dallas, TX 75275

SUMMARY

It has been suggested that line transect methods should be "pooling robust" (Buckland et al, 1993;Burnham et al, 1980) allowing for transect data from different transects or locations to be pooled for estimation of population density. This is particularly important in situations where data from individual transects are sparse and pooling is done out of necessity. In this study we investigate a method for combining estimates from individual transects when each transect has sufficient data to support estimation with the kernel method. It is based on a minimizer of the asymptotic mean squared error of a convex linear combination of the individual population density estimators. It is shown that the asymptotic mean squared error of the simple pooled estimator is always at least as large as the optimally combined estimator. We provide an application that combines two estimates for a real population of mussels. Using a variety of simulations, we demonstrate the better finite sample efficiency for combining unbalanced cases. When the detection functions are identical, we find it better to pool.

key words:ecology,nonparametric,survey

1. Introduction

Line transect sampling methods are commonly used by biologists to estimate population density. In line transect sampling, distances from selected transect lines are used to construct an estimate of population density which involves estimating the probability density of sighting distances on the transect line (distance = 0). Recent work has focused on employing kernel methods to estimate the probability density (Mack and Quang, 1998; Chen, 1996). This method does not require a parametric form of the probability density to be specified, but does require specification of the amount of smoothing to be done through a smoothing parameter or bandwidth.

It has been suggested by some authors (Buckland et al, 1993; Burnham et al, 1980) that line transect estimators be "pooling robust." That is, the estimators should yield consistent estimates of population density when distances from different transects or groups of transects are pooled and treated as one dataset. This quality is especially pertinent in those situations where individual transects or naturally grouped transects do not yield sufficient sightings to provide adequate individual estimates of population density. Hence the data are pooled out of necessity. In situations where individual transects, or groups of transects, have adequate data to form individual estimates of population density, more efficient estimates may be formed by combining these estimators in some way other than simple pooling.

In this paper we investigate an alternative method of combining kernel based population density estimates from different sources under the assumption that the underlying true population density is constant. The estimator is based on the minimizer of the asymptotic mean square error of a linear combination of the individual population density estimates and is similar in nature to the estimators studied in Gerard and Schucany (1996,1997) for combining nonparametric regression estimates. Different transect line lengths can be accommodated as can different detection functions, so long as the probability of detection on the transect line is 1.0. Differences

2

in detection functions could be caused by a variety of factors including differences in observers or weather conditions.

Consider distance data for a common species of mussels from two transects. The first transect was 42 meters in length and yielded distances for 53 mussels while the second was 52 meters in length and yielded 233 sighting distances. The data for each transect are displayed along the horizontal axis in Figure 1a and 1b, respectively, with the corresponding kernel density estimate of sighting distances. Each kernel estimate uses a quadratic weight function and global bandwidths of 25 and 15 meters, respectively. The goal is to combine the individual estimators based on these kernel estimators of probability density to form a single estimate of population density.

[Figure 1 here]

In Section 2 we will review kernel estimation of population density and describe the two estimators to be compared. We then evaluate the asymptotic relative efficiency of the optimally combined estimator to the pooled estimator by comparing there asymptotic mean squared errors. In Section 3, these two estimators are compared using Monte Carlo simulation techniques. We will revisit the example introduced in the previous paragraph in Section 4 and then provide some discussion in Section 5. Proof of a theorem concerning the dominance of the optimally combined estimator over the simple pooled estimator is provided in a technical appendix.

2. Kernel Estimation of Population Density

The typical estimate of population density is

$$\hat{D} = n\hat{f}(0)/(2L),$$ (2.1)

where n is the number of objects detected, L is the length of the transect line, and $\hat{f}(0)$ is an estimate of the probability density of sighting distances on the line. For kernel estimates,
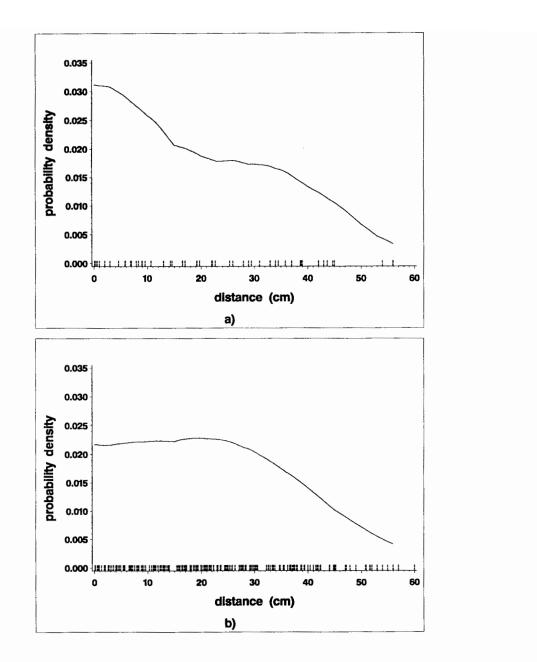
3

Figure 1. Estimated Kernel Densities of Sighting Distances of Mussel Line
Transect Sightings using global bandwidths of a) 25 meters and b) 15 meters and
the quadratic weight function.

$\hat{f}(0) = n^{-1} \sum_{i=1}^{n} K_h^*(x_i)$, where h is a bandwidth that governs the amount of smoothness in the

estimators, x_i is the ith distance from the object to the transect line, and $K_h^*(u) = h^{-1} K^*(u/h)$ is a

kernel function adapted for estimation at the edge of the estimation space ($K^*(\cdot) = 2K(\cdot)$), where

$K(\cdot)$ is a kernel function as defined in Silverman (1986) or Wand and Jones (1995).

In most instances, distance data is available from more than one transect or group of

transects. Hence, the simple pooled estimator, $\hat{D}_p$, in the case where two transects are available

is

$$\hat{D}_p = \sum_{i=1}^{2} \sum_{j=1}^{n_i} K_h^*(x_{ij}) \Big/ \{2(L_1 + L_2)\}, \tag{2.2}$$

where n_i is the number of sightings for the ith transect, i=1,2; x_{ij} is the distance for the jth sighting

in the ith group, j=1 .. n_i; and L_i is the transect length for the ith transect. Hence if we let $\hat{D}_i$ be

the ith transect estimate as in (2.1), then $\hat{D}_p = p_L \hat{D}_1 + (1 - p_L)\hat{D}_2$, where $p_L = L_1 / (L_1 + L_2)$.

Note that the pooled estimator is a linear combination of the individual density estimates for each

transect. The weighting factor involves only the length of the transect, with the longer transect

receiving more weight. If the detection function differs for the two transects, this estimator is not

able to weight the individual estimator to account for this difference.

Evaluation of $\hat{D}_p$ and the estimator to be proposed is facilitated if it is noted that the

numerator of (2.1) is equivalent in form to an estimator of a varying Poisson intensity, λ,

evaluated on the transect line (Diggle and Marron, 1985). As noted in Cowling, Hall, and

Phillips (1996), consistent estimation of λ by $\hat{\lambda} = n\hat{f}(0)$ employing kernel methods requires that

λ increases. Hence asymptotic arguments typically make use of the assumption that $\lambda = \mu r$, say,

4

where μ is a constant function and $r \to \infty$. Hence in our application it is reasonable to assume that $\lambda = 2L\mu$ with μ being the true underlying density to be estimated with $\hat{D}$.

It is straightforward to show using the results of Cowling, Hall, and Phillips (1996) that the asymptotic mean squared error is

$$AMSE\left[\hat{D}_p\right] = (1/4)k_2^{*2}h^4\left\{p_L\mu_1'' + (1-p_L)\mu_2''\right\}^2 + \mu Q^*/\left\{2(L_1+L_2)h\right\}, \tag{2.3}$$

where $k_2^* = \int_0^\infty u^2 K^*(u)du$ and $Q^* = \int_0^\infty K^{*2}(u)du$. The first term in (2.3) is the squared bias and

the second is the variance. Although the underlying density, μ, is assumed to be the same for

each transect line, differing detection functions result in different second derivatives, μ_i''. The

solution to the derivative of (2.3) with respect to h yields an asymptotically optimal bandwidth

and substitution into (2.3) results in the minimum AMSE for $\hat{D}_p$, which will be used in

Asymptotic Relative Efficiency (ARE) evaluations done subsequently.

Suppose we now consider a general convex linear combination $\hat{D}_c = c\hat{D}_1 + (1-c)\hat{D}_2$

($0<c<1$). The asymptotic AMSE is

$$AMSE\left(\hat{D}_c\right) = (1/4)\left\{c\left(\mu_1''h_1^2k_2^*\right) + (1-c)\left(\mu_2''h_2^2k_2^*\right)\right\}^2 + c^2\mu Q^*/(2L_1h_1) + (1-c)^2\mu Q^*/(2L_2h_2). \tag{2.4}$$

Define $\mu_c = \sqrt{\mu_2''/\mu_1''}\,\mu$ and $h_{2c} = \sqrt{\mu_2''/\mu_1''}\,h_2$ and assume that the second derivatives have the

same sign. Hence, $\mu_c = \mu|\mu_2''|^{1/2}/|\mu_1''|^{1/2}$ and $h_{2c} = h_2|\mu_2''|^{1/2}/|\mu_1''|^{1/2}$. The AMSE in (2.4)

can be expressed as

$$AMSE\left(\hat{D}_c\right) = (1/4)\left\{c\left(\mu_1''h_1^2k_2^*\right) + (1-c)\left(\mu_2''h_{2c}^2k_2^*\right)\right\}^2 + c^2\mu Q^*/(2L_1h_1) + (1-c)^2\mu_c Q^*/(2L_2h_2). \tag{2.5}$$

5

This is the same general form of Theorem 1 in Gerard and Schucany (1996). Hence, the values of c, h_1, and h_{2c} that minimize (2.5) are

$$\tilde{c} = (\mu_c/L_2)/(\mu/L_1 + \mu_c/L_2) = \left(|\mu_2''|^{1/2}/L_2\right)\Big/\left(|\mu_1''|^{1/2}/L_1 + |\mu_2''|^{1/2}/L_2\right) \qquad (2.6)$$

and

$$\tilde{h}_1 = \tilde{h}_{2c} = \left[\left(|\mu_1''|^{1/2}/2L_1\right)\left(|\mu_2''|^{1/2}/2L_2\right)\mu Q^* \Big/ \left\{\left(|\mu_1''|^{1/2}/2L_1 + |\mu_2''|^{1/2}/2L_2\right)|\mu_1''|^{1/2}\left(k_2^* \mu_1''\right)^2\right\}\right]^{1/5}.$$

$$(2.7)$$

This leads to $\tilde{h}_2 = \tilde{h}_{2c}k_d$, where

$$k_d = |\mu_1''|^{1/2}\Big/|\mu_2''|^{1/2}. \qquad (2.8)$$

Using these values yields $\hat{D}_{\tilde{c}}$. It should be noted that if $\mu_1'' = \mu_2''$, then this estimator is asymptotically equivalent to $\hat{D}_p$. However, if the second derivatives differ, $\hat{D}_{\tilde{c}}$ weights the individual density estimates accordingly. For example for equal transect lengths, if μ_2'' is large compared to μ_1'' the population density estimate for the first transect is weighted more heavily. Hence, this estimator attempts to simultaneously correct for differences in transect length and detection function. It should also be noted that the bandwidths used in (2.7) yield density estimates for the two transects with equal bias. Proof that this solution results in a local minimum for (2.5) follows from Theorem 1 of Gerard and Schucany (1996).

It can be shown that $AMSE(\hat{D}_{\tilde{c}}) \le AMSE(\hat{D}_p)$ (Theorem 1 in Appendix). A contour plot of the asymptotic relative efficiency (see page 106, Wand and Jones, 1995) of $\hat{D}_{\tilde{c}}$ relative to $\hat{D}_p$, $\left\{AMSE(\hat{D}_p)\big/AMSE(\hat{D}_{\tilde{c}})\right\}^{5/4}$, as a function of k_d in (2.8) and $k_L = L_1/L_2$ is shown in Figure 2. This plot provides contours of equal values of ARE in increments of .2. It is evident from the

6

plot that the effect due to differences in transect lengths is not as pronounced as the effect due to differences in detection functions. For most values of k_L, values of k_d between .4 and 1 yield ARE values less than 1.20. Values of k_d less than .2 yield ARE values that increase sharply as k_d decreases. Of course in practice the weighting factors and bandwidths involve quantities that are unknown and must be estimated. Hence, when adequate sample sizes are available for each individual estimate of population density to be estimated, there is potential for much greater efficiency when optimally combining rather than pooling. In Section 3 we compare the finite sample properties of these two estimators using Monte Carlo simulation.

[Figure 2 here]

3. A Monte Carlo Comparison of $\hat{D}_p$ and $\hat{D}_{\tilde{c}}$

In order to evaluate the finite sample performance of these two estimators, a simulation study was conducted. The exponential power function, $g(x) = \exp\left\{-(bx)^a\right\}$, and a detection function based on the t distribution, $g(x) = \left(1 + x^2/k\right)^{-(k+1)/2}$ with $k = 4$, were used. A constant density of .15 was used in all simulation runs, with the length of the transect line and the number of objects placed in the study area altered to achieve this density. The effective width of the transect area was taken to be 10. The distance from the transect line for each object was determined from a uniform random variable in the interval [0,10]. Objects were considered observed if the probability of detection, $g(x)$, at the generated distance was greater than a uniform random variable in the interval [0,1]. Distances for two independent transect lines were generated, possibly with different detection functions and transect lengths. Five thousand replications were generated under each set of conditions.

For each replication, both $\hat{D}_p$ and $\hat{D}_{\tilde{c}}$ were computed. In order to compute $\hat{D}_p$ as in (2.2), a single bandwidth needs to be specified. The bandwidth used was a local normal scale rule that
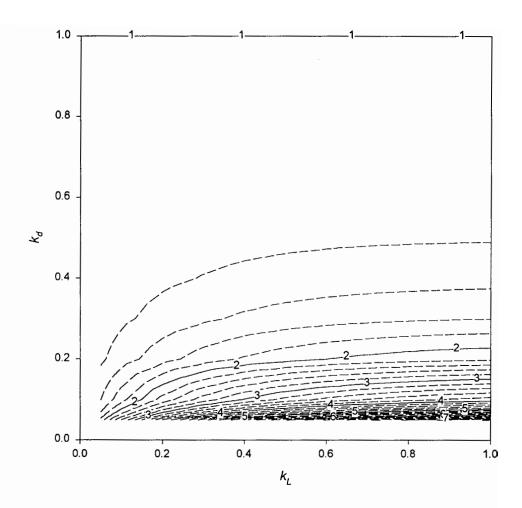
7

Figure 2. Asymptotic Relative Efficiency of $\hat{D}_{\tilde{c}}$ relative to $\hat{D}_p$.

has been found to be as effective or more so than some sophisticated local plug-in rules (Gerard and Schucany, 1999). Essentially, the unknown quantities in the asymptotically optimal bandwidth from (2.3) are estimated by treating the underlying probability distribution of the sighting distances as the absolute value of a normal variable with mean zero. Hence, we assume that f that is estimated in the numerator of (2.1) is that of the absolute value of a normal variable with mean zero. Thus estimation amounts to estimating σ^2 by $\hat{\sigma}^2 = \sum_{i=1}^{n} x_i^2 \Big/ n$, where the x_i are the distances from the transect line, and inserting this estimate into either f or f_i'' . To compute $\hat{D}_{\tilde{c}}$, two bandwidths in (2.7) and a weighting factor in (2.6) must be estimated. Again, a normal scale rule will be used to estimate the unknown quantities μ and μ_i'' . Estimation of μ and μ_i'' is essentially equivalent to estimation of f and f_i'' , respectively. Because we have two sources to use when estimating μ in (2.7), the normal scale estimates from each source are averaged. Additionally, it should be noted that use of the normal scale rule assures us that our estimates of μ_i'' will have the same sign, as required for our asymptotic results.

The results of the simulation study are given in Table 1. For situations where the estimators are asymptotically equivalent, where the detection functions are the same for each transect, $\hat{D}_p$ performs better. This is likely to be due to the added variability caused by estimation of the weighting factor in (2.6). In those cases where the detection functions are quite different, $\hat{D}_{\tilde{c}}$ tends to perform better.

[Table 1 here]

In the next section we will use these two methods to calculate the population density of mussels.

4. Application to estimation of mussel population density

8

Table 1

Summary of Simulation Averages for Optimal and Pooled Estimators Based on 5000 Replications

Transect 1			Transect 2			MSE x 10^4		
g^1	L	n	g	L	n	$\hat{D}_c$	$\hat{D}_p$	RE^2
1	133	71	1	133	71	3.84	3.66	.94
1	133	71	1	67	35	4.76	4.53	.94
1	133	71	2	67	35	4.61	4.37	.94
2	67	36	2	67	36	6.08	5.84	.95
1	133	71	3	133	53	5.53	4.77	.83
4	133	35	1	133	71	4.97	5.81	1.22
4	133	35	4	67	18	9.06	8.29	.89
5	133	142	4	133	36	3.00	18.8	9.91
5	100	106	1	100	53	3.20	4.67	1.60
3	133	53	5	100	106	4.35	7.87	2.10

[1] detection function, g =1 (Power(b=.5,a=2)), 2 (Power(b=.5,a=2.5)), 3 (t 4df), 4 (Power(b=1,a=1), 5 (Power(b=.25,a=2)).

[2] Estimated Relative Efficiency, $RE = \left\{ MSE\left(\hat{D}_p\right) \middle/ MSE\left(\hat{D}_{\tilde{c}}\right) \right\}^{5/4}$

Consider the data from two transects displayed in Figure 1. Using a normal scale rule as described in Section 3 for bandwidth selection, $\hat{D}_p$ yields an estimate of population density of .000352/cm². Individual density estimates using the bandwidths as in (2.7) estimated analogously as in Section 3 were .000167/cm² using a bandwidth of 25 meters for the first transect and .000486 using a bandwidth of 15 meters for the second transect. Optimally combining them yields $\hat{D}_{\tilde{c}} = .000300/\text{cm}^2$. The pooled estimator weighs the individual estimators in a very similar fashion because the transect lengths are nearly the same (42 meters and 52 meters, respectively). The optimally combined estimator weighs the second individual estimator more heavily because there is less curvature in the probability density of sighting distances on the line and therefore a smaller second derivative.

5. Discussion

In this paper we have presented an alternative estimator to the usual simple pooled estimator that can be used in situations where sufficient sightings are available for each transect to reasonable compute individual estimates of population density. It adapts to differences in transect length as well as differences in detection functions for the two transects. Simulation studies show that in situations where asymptotic results would suggest that the pooled estimator and the optimally combined estimator should perform similarly, the simple pooled estimator typically performs better. This is probably due to increased variability stemming from estimation of second derivatives required in the weighting factor and bandwidths. In those situations where the detection functions are quite different, gains from using the optimally combined estimator are evident. Extension to situations with more than two transect lines should be relatively straightforward.

In addition to the application of our proposed estimator to estimation of population density from two transects when the underlying population density is assumed to be the same for each

transect, it also has potential use in hypothesis testing. Mack and Quang (1998) showed that estimates of population density using the kernel method are asymptotically normal. It may be of interest to use kernel based estimates of population density to test whether population density differs for two locations in a study area, for example. It may not be reasonable to assume that the detection function is the same for each area. An estimate of the variance of the difference between independent kernel estimates of population density under the null hypothesis of equal density requires that one compute an estimate of density under that assumption. Our proposed estimator does just that. It has the added benefit of providing bandwidths that yield individual estimates of equal bias. Equal bias simplifies testing because the numerator of the test statistic involves forming a difference between the estimated population densities for the two locations, which essentially cancels the bias terms.

Acknowledgements

Appendix

Theorem 1

If the signs of μ_1'' and μ_2'' are the same, then $AMSE\left(\hat{D}_p\right) \geq AMSE\left(\hat{D}_{\tilde{c}}\right)$, with equality when $k_d = 1$.

Proof:

We can express $AMSE\left(\hat{D}_p\right) = AMSE\left(\hat{D}_2\right)\left\{\left(k_d^2 k_L + 1\right) / \left(k_L + 1\right)^3\right\}^{2/5}$, where k_d is as in (2.8) and $k_L = L_1 / L_2$. Similarly, the asymptotic mean squared error of $\hat{D}_{\tilde{c}}$ can be expressed

10

as $AMSE\left(\hat{D}_{\tilde{c}}\right) = AMSE\left(\hat{D}_2\right)\{k_d/(k_L + k_d)\}^{4/5}$. Form the ratio $AMSE\left(\hat{D}_p\right)/AMSE\left(\hat{D}_{\tilde{c}}\right) =$

$\left[\left(k_d^2 k_L + 1\right)\left(k_d + k_L\right)^2/\left\{\left(k_L + 1\right)^3 k_d^2\right\}\right]^{2/5}$. This reduces to

$\left\{\left(k_d^2 k_L^3 + k_d^2 + 2k_L^2 k_d^3 + k_d^4 k_L + k_L^2 + 2k_L k_d\right)/\left(k_d^2 k_L^3 + k_d^2 + 3k_L^2 k_d^2 + 3k_L k_d^2\right)\right\}^{2/5}$. Note that the

first two terms are the same in both numerator and denominator. Additionally,

$2k_L^2 k_d^3 + k_L^2 \geq 3k_L^2 k_d^2$, with equality when k_d=1 or k_L=0. Finally, $k_d^4 k_L + 2k_L k_d \geq 3k_L k_d^2$,

again with equality if k_d=1, k_d=0, or k_L=0. It follows that the ratio is greater than or equal to

unity.‖

References

Buckland, S.T., Anderson, D.R., Burnham, K.P., and Laake, J.L. (1993). *Distance Sampling*. London : Chapman and Hall.

Burnham, K.P., Anderson, D.R., and Laake, J.L. (1980). *Estimation of Density from Line Transect Sampling of Biological Populations*. Wildlife Mongraph Number 72.

Chen S.X. (1996). Studying school size effects in line transect sampling using the kernel method. *Biometrics* 52, 1283-1294.

Cowling, A., Hall, P., and Phillips, M.J. (1996). Bootstrap confidence intervals for the intensity of a Poisson point process. *Journal of the American Statistical Association* 91, 1516-1524.

Diggle, P. and Marron, J.S. (1988). Equivalence of smoothing parameter selectors in density and intensity estimation. *Journal of the American Statistical Association* 83, 793-800.

Gerard, P.D. and Schucany, W.R. (1996). On combining independent nonparametric regression estimators. *Statistics and Probability Letters* 26, 25-31.

Gerard, P.D. and Schucany, W.R. (1997). Methodology for nonparametric regression from independent sources. *Computational Statistics and Data Analysis* 25, 287-304.

Gerard, P.D. and Schucany, W.R. (1999). Local bandwidth selection for kernel estimation of population densities with line transect sampling. *Biometrics*. To Appear.

Mack, Y.P. and Quang, P.X. (1998). Kernel methods in line and point transect sampling. *Biometrics* 54, 606-619.

Silverman, B.W. (1986). *Density Estimation*. London : Chapman and Hall.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. London : Chapman and Hall.