Kernel Methods for Estimating the Utilization Distribution in Home-Range Studies

B. J. Worton

Ecology, Volume 70, Issue 1 (Feb., 1989), 164-168.

Stable URL:
http://links.jstor.org/sici?sici=0012-9658%28198902%2970%3A1%3C164%3AKMFETU%3E2.0.CO%3B2-Y

Ecology, 70(1), 1989, pp. 164–168

KERNEL METHODS FOR ESTIMATING THE UTILIZATION DISTRIBUTION IN HOME-RANGE STUDIES[1]

B. J. WORTON
Mathematical Institute, University of Kent, Canterbury, Kent CT2 7NF England

Abstract. In this paper kernel methods for the nonparametric estimation of the utilization distribution from a random sample of locational observations made on an animal in its home range are described. They are of flexible form, thus can be used where simple parametric models are found to be inappropriate or difficult to specify. Two examples are given to illustrate the fixed and adaptive kernel approaches in data analysis and to compare the methods. Various choices for the smoothing parameter used in kernel methods are discussed. Since kernel methods give alternative approaches to the Anderson (1982) Fourier transform method, some comparisons are made.

Key words: bivariate density estimation; data analysis; home range; kernel estimate; nonparametric estimation; probabilistic model; utilization distribution.

INTRODUCTION

Estimation of the utilization distribution (UD: the name given to the distribution of an animal's position in the plane) is of great importance in home-range studies. Both parametric (Jennrich and Turner 1969, Van Winkle 1975, Don and Rennolls 1983) and nonparametric (Ford and Krumme 1979, Dixon and Chapman 1980, Anderson 1982) methods have been used to estimate the UD. The purpose of this paper is to illustrate how one might proceed with home-range data analysis if simple parametric models are inappropriate for the UD.

Many papers have been published on the subject of home-range estimation and reviews are given in, for example, Macdonald et al. (1980) and Worton (1987). Although approaches such as polygon and grid cell methods have been used for data analysis, here we concentrate on probability density estimation approaches. That is, the home range of an animal is described in terms of a probabilistic model. If an estimate of home range is required from the estimated UD, the area contained within the $100\alpha\%$ confidence region may be used. Values of $\alpha = .5$ and .95 are commonly used, but there is possibly some advantage in using the smaller value of $\alpha = .5$ (see Anderson 1982). This paper is not intended to review the home-range literature, but it is useful to describe some approaches that have been used to estimate the UD.

The first methods used to estimate UDs assumed simple bivariate normal models (Calhoun and Casby 1958, Jennrich and Turner 1969). More recently Don and Rennolls (1983) used a mixture of bivariate normal distributions that allowed for the possibility of multimodality. It is possible to think of other, more sophisticated, models that could be used to describe the UD. However, it is worth keeping models as simple as possible, but still having a reasonable fit for the data. Because of the processes that give rise to home range area usage, it is not difficult to imagine that UDs might arise that would be difficult to model using standard bivariate distributions (Dixon and Chapman 1980, Anderson 1982). For this reason, when a simple parametric model is inappropriate or difficult to specify, there is a need for nonparametric estimation approaches.

The harmonic mean method proposed by Dixon and Chapman (1980) describes the intensity of use of the home range at a point $\boldsymbol{x}$ by

$$HM(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} |\boldsymbol{x} - \boldsymbol{x}_i|^{-1},$$

where $\boldsymbol{x}_i$, for $i = 1, \ldots, n$, are the observational data. Unfortunately, since the method does not produce a probability density, it has only a limited probabilistic interpretation. The approach also has the disadvantage of producing misleading results if an observation is near $\boldsymbol{x}$, the point of evaluation. Various modifications to the method have been made to overcome these problems (Dixon and Chapman 1980, Spencer and Barrett 1984, Samuel et al. 1985).

Anderson (1982) discussed how bivariate histograms may be used to estimate the UD. However, histogram estimates are very subjective because one has to choose (1) the origin, (2) the cell size, and (3) the orientation of the axes. To avoid these problems and the inefficient use of data, Anderson used the Fourier transform method to estimate the UD (see Tarter and Kronmal 1970). Essentially, the method estimates the density by a series of sines and cosines of different frequencies. High-frequency terms in the series can be removed to leave a "smoothed" density. Anderson modified the original method of Tarter and Kronmal that tests if terms in the series contribute significantly to avoid unacceptable, jagged density estimates.

Two possible disadvantages with the Fourier trans-

[1] Manuscript received 25 January 1988; revised and accepted 25 May 1988.

form method are that the density estimates can take negative values, and estimation can only be made on a specified finite region of the plane. Kernel methods, as described below, are free of these problems and provide alternative approaches. They are also conceptually easier to understand and explain. It is not suggested that kernel methods are better than the Fourier transform method, although there are reasons why biologists may prefer to use them in home-range data analysis.

Kernel Methods

Like the Fourier transform method, kernel methods free the UD estimate from parametric assumptions and provide a means of smoothing locational data to make more efficient use of them than a histogram. Kernel methods have well-understood consistent statistical properties (Rosenblatt 1956, Parzen 1962, Cacoullos 1966) and are widely used in both univariate and multivariate probability density estimation.

Fixed kernel method

Restricting ourselves to the bivariate case, suppose that $\boldsymbol{X}_1 = [X_1^{(1)}, X_1^{(2)}]'$; $\boldsymbol{X}_2 = [X_2^{(1)}, X_2^{(2)}]'$; . . . , $\boldsymbol{X}_n = [X_n^{(1)}, X_n^{(2)}]'$ is a random sample of n independent points from an unknown utilization distribution with probability density function $f(\boldsymbol{x})$, which is to be estimated, then the bivariate kernel estimator of $f(\boldsymbol{x})$ can be defined as

$$\hat{f}_h(\boldsymbol{x}) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left[\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h}\right],$$

where the kernel K is a unimodal symmetrical bivariate probability density function, and h is the smoothing parameter that can be varied by the user (Silverman 1986). As an example, if the bivariate normal density kernel is used, the kernel density estimator is

$$\hat{f}_h(\boldsymbol{x}) = \frac{1}{nh^2} \sum_{i=1}^{n} \frac{1}{2\pi} \cdot \exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{X}_i)'(\boldsymbol{x} - \boldsymbol{X}_i)}{2h^2}\right).$$

Alternatively, the Epanechnikov kernel defined as

$$K(\boldsymbol{x}) = \begin{cases} 2\pi^{-1}(1 - \boldsymbol{x}'\boldsymbol{x}) & \text{for } \boldsymbol{x}'\boldsymbol{x} < 1 \\ 0 & \text{otherwise} \end{cases}$$

may be used to obtain the kernel estimator. At this stage it is worth describing an intuitive interpretation of the kernel method.

The kernel estimator may be pictured as follows. A scaled-down probability density function, namely the kernel, is placed over each data point and the estimator is constructed by adding the n components. Thus, where there is a concentration of points the kernel estimate has a higher density than where there are few points. Because each kernel is a density the resulting estimate is a true probability density function itself.

The smoothing parameter controls the amount of variation in each component of the estimate. Thus, if a small value of h is used, the fine detail of the data can be observed, while a larger value of h obscures all but the most prominent features. The type of kernel estimators considered so far are called fixed kernel estimators because the smoothing parameters are of fixed value over the plane.

It is interesting to note that the harmonic mean method can be interpreted as a kernel method with kernel

$$K(\boldsymbol{x}) = |\boldsymbol{x}|^{-1}.$$

However, since this type of kernel does not have the properties of a probability density function and there is no smoothing parameter used, the method does not possess the desirable properties of the usual fixed kernel method.

Adaptive kernel method

Sometimes a more sophisticated kernel approach called the adaptive kernel method is needed. This method varies the smoothing parameter so that areas with a low concentration of points have higher h values than areas with a high concentration of points, and are thus smoothed more. Therefore, the adaptive method is an improvement on the fixed kernel method, particularly in the tails of the density. The UD density estimator is

$$\hat{f}_h(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i^2} K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h_i}\right),$$

where the smoothing parameters h_i are based on some "pilot" estimate of the density. Silverman (1986:101) suggested that the smoothing parameters may be taken as $h_i = h\lambda_i$ where h is a global smoothing parameter and the λ_i are local smoothing parameters given by

$$\lambda_i = \left\{\frac{\tilde{f}(\boldsymbol{X}_i)}{g}\right\}^{-1/2} \quad \text{for } i = 1, \ldots, n,$$

where

$$\log g = \frac{1}{n} \sum_{i=1}^{n} \log \tilde{f}(\boldsymbol{X}_i)$$

and $\tilde{f}(\boldsymbol{X}_i) > 0$ is some pilot estimate of $f(\boldsymbol{x})$ evaluated at point $\boldsymbol{X}_i$. Adopting this approach there is only one smoothing parameter, as with the fixed kernel method. As an example, the pilot estimate $\tilde{f}(\boldsymbol{x})$ may be taken as a fixed kernel estimate. Breiman et al. (1977) considered another form of this type of estimator.

Theoretical optimum choice of smoothing parameter

It is generally accepted in the kernel literature that the choice of the kernel K is not as important as the

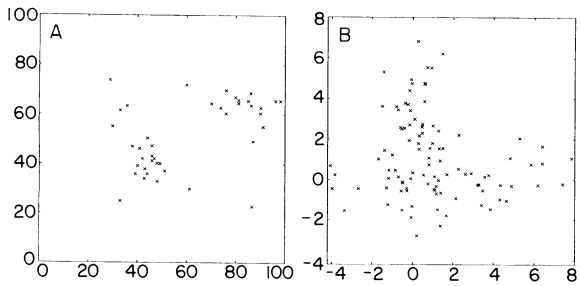

FIG. 1. Plots of (A) the DC data set and (B) the SIM data set.

choice of the smoothing parameter (Epanechnikov 1969). Thus, for a given kernel and sample size we need to find the "best" value of h. The criterion for choosing a good value of h is often given in terms of the global measure of error, namely the mean integrated square error ($MISE$) defined as

$$MISE(h) = E \int (\hat{f}_h - f)^2,$$

where E denotes the expectation with respect to the random observations. For our bivariate case, the integration is over the plane. The best value of h is taken as the value of h that minimizes $MISE(h)$.

Unfortunately, from our practical interest in the choice of h, some theoretical calculations show that to obtain the optimum smoothing parameter, using this criterion, we need to know the UD density $f(\boldsymbol{x})$. Therefore, h cannot be obtained in this way in practice. Some possible approaches for choosing h in data analysis are discussed in the following section.

Choosing the smoothing parameter in practice

An ad hoc, but popular, method for choosing h is to use the optimum h value obtained for some standard distribution, such as the normal distribution. Using a fixed normal kernel, it can be shown that for the bivariate normal distribution with variance–covariance matrix $\begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$, the optimum h for large sample size n is

$$h_{\text{opt}} = \sigma n^{-1/6}.$$

Thus, an obvious estimate of h_{opt} is

$$\hat{h}_{\text{opt}} = \hat{\sigma} n^{-1/6},$$

where $\hat{\sigma} = \{1/2\ [\hat{\sigma}^2_{x(1)} + \hat{\sigma}^2_{x(2)}]\}^{1/2}$, and $\hat{\sigma}^2_{x(1)}$ and $\hat{\sigma}^2_{x(2)}$ are the estimated variances of the $x^{(1)}$ and $x^{(2)}$ data. If these variances differ greatly it may be worth rescaling the data so that the variances are equal before applying a kernel method. Bowman (1985) has shown that this estimation method often produces better results than more sophisticated methods in univariate estimation. However, caution should be exercised in using this choice of h, especially if the UD is suspected to be multimodal.

Of the objective methods for estimating h described in Silverman (1986), the approach of least-squares cross-validation appears to provide a good method of estimation. If a standard bivariate normal density fixed kernel K is used, the value of h is chosen to minimize

$$M(h) = n^{-2}h^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} K^*\left(\frac{\boldsymbol{X}_i - \boldsymbol{X}_j}{h}\right) + 2n^{-1}h^{-2}K(\boldsymbol{0}),$$

where $K^* = K^{(2)} - 2K$, and $K^{(2)}$ is the bivariate normal density with variance–covariance matrix $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. The intuitive reason this method provides a good estimate is that

$$E[M(h)] + \int f^2 \approx E \int (\hat{f}_h - f)^2.$$

Thus, by minimizing $M(h)$ we would hope to minimize $MISE(h)$. Least-squares cross-validation can also be used to find the global smoothing parameter in the adaptive kernel method.

Home-Range Data Analyses

Two data sets are analyzed in this section to illustrate the practical use of kernel methods in home-range data analysis. The first data set is taken from Dixon and Chapman (1980:Fig. 3) with sample size $n = 41$, and the second is an artificial data set of sample size $n = 100$ from a bivariate distribution with nonconvex probability density function contours; these data sets are denoted by DC data and SIM data respectively. The bivariate distribution from which the artificial data were simulated has the probability density function

$$f(x, y) = \frac{1}{12\pi}\left\{\exp\left[-\frac{(x - 2.5)^2}{18} - \frac{y^2}{2}\right] + \exp\left[-\frac{x^2}{2} - \frac{(y - 2.5)^2}{18}\right]\right\}.$$

This is a mixture of normal densities.

By simply inspecting scatter plots of the two data sets (Fig. 1) it may not be clear what type of parametric

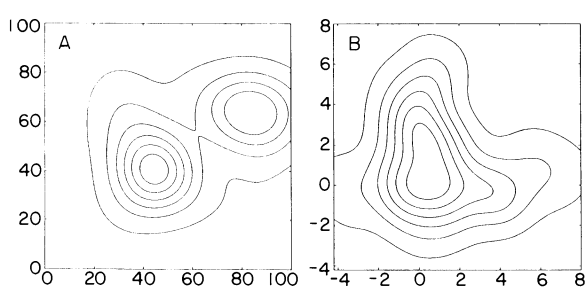

FIG. 2. Fixed kernel density estimates of the UD densities with the ad hoc choice of smoothing parameters for (A) the DC data set ($h = 10.$) and (B) the SIM data set ($h = 1.$).

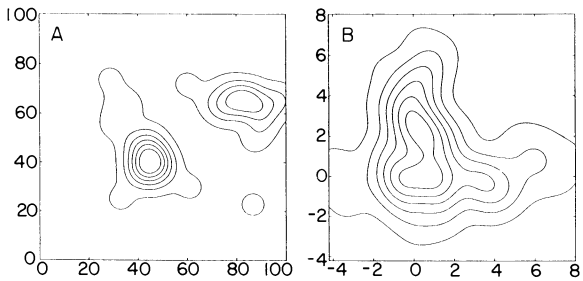

FIG. 3. Fixed kernel density estimates of the UD densities with the least-squares cross-validation choice of smoothing parameters for (A) the DC data set ($h = 4.7$) and (B) the SIM data set ($h = 0.77$).

model would be appropriate. Clearly, the bivariate normal parametric model, which has been widely used in past home-range analyses, is not a plausible model for either data set. Since both DC and SIM types of data appear to occur frequently, their analyses are of particular interest.

The standard normal kernel is used below in both fixed and adaptive kernel estimates.

DC data analysis

First, consider fixed kernel estimation of the UD density. As an exploratory exercise, the UD was estimated for a range of h values by the fixed kernel method. This gave a feel for the nature of the data and showed the two-cluster character of the data more clearly than in the scatter plot. Using the ad hoc estimate of h we find that $h = 10$ and produces a UD density estimate illustrated in Fig. 2A. Using the method of least-squares cross-validation, the score function $M(h)$ gave a minimum value at $h = 4.7$, and Fig. 3A is an appropriate UD density estimate.

Comparing the two results we see that the least-squares cross-validation choice of h gives a clearer picture of the underlying UD density, while the ad hoc choice of h has over-smoothed the data. However, both show that the UD is clearly not unimodal. The least-squares cross-validation choice could also be suggesting a third important mode above the left cluster.

The bump in the fixed kernel density estimate in Fig. 3A at the bottom right of the plot is due to a point in an area with a low concentration of points. To remove this effect, an adaptive kernel estimate was constructed, with a least-squares cross-validation choice of $h = 5.7$. This is illustrated in Fig. 4A, and the pilot estimate is taken as the fixed kernel estimate with $h = 4.7$. The adaptive estimate shows very clearly the bimodal nature of the data, and gives a better picture of the tails of the UD than the fixed kernel estimates.

Very similar estimates indeed were obtained when the above analyses were carried out using the Epanechnikov kernel. This supports the theoretical findings that the precise form of the kernel used is unimportant.

SIM data analysis

For the fixed kernel method, the ad hoc choice of h is 1, and the least-squares cross-validation choice of h is 0.77. The respective UD kernel density estimates are shown in Figs. 2B and 3B. As with the DC data, the ad hoc choice produces an over-smoothed estimate compared with the least-squares cross-validation choice.

The adaptive kernel estimate based on a least-squares cross-validation choice of $h = 0.79$ produces the estimate shown in Fig. 4B. Comparing Figs. 3B and 4B we see that the latter estimate gives a more satisfactory representation of the tails of the UD density.

Although the above three kernel estimates all show the important nonconvex feature of the UD density, which is illustrated for the SIM data in Fig. 5, the adaptive kernel estimate gives the most accurate picture.

Discussion and Conclusions

The adaptive kernel estimate with a least-squares cross-validation choice of h produced the best results for both DC and SIM data sets, but the fixed kernel method with an ad hoc choice of h also produced estimates of practical value. Generally the kernel method and choice of smoothing parameter method adopted in home-range data analysis will depend on the intended use of the UD density estimate. If accuracy is not crucial, the fixed kernel method with ad hoc choice of h may provide all the information one needs about the UD, for example, the location of the modes. However, if one is investigating UDs, say, for static attraction/repulsion, then accuracy in the tails of the density estimates is important, and the adaptive method with least-squares cross-validation choice of h should be used.

Computationally, fixed kernel estimates are the same to evaluate as the harmonic mean method. The adaptive method needs more calculation since the pilot estimate on which the local smoothing parameters are based has to be obtained first. Therefore, as one would

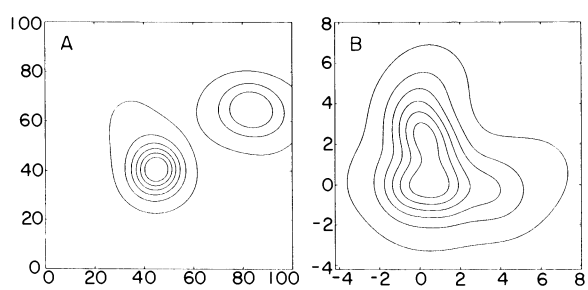

FIG. 4. Adaptive kernel density estimates of the UD densities with the least-squares cross-validation choice of smoothing parameters for (A) the DC data set ($h = 5.7$) and (B) the SIM data set ($h = 0.79$). The pilot estimates were taken as the estimates shown in Fig. 3.

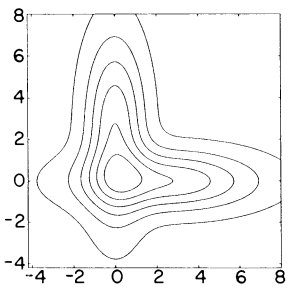

FIG. 5. Contours of the true probability density function from which SIM data were simulated.

expect, more computation is needed for the better kernel estimate.

The above analyses show that kernel methods are very useful for estimating the UD density because there are no constraints placed on the form of the UD. However, the nonparametric kernel approaches should not be used to the exclusion of parametric approaches, and if a suitable parametric model is available there is no reason why it should not be used. At the exploratory stage of data analysis, a kernel estimate with a subjectively chosen smoothing parameter may help to suggest a suitable parametric model for the UD. Thus, parametric and nonparametric methods can be used to investigate home-range data, but nonparametric approaches have the great advantage of flexibility. Often, only a kernel estimate of the UD density will be needed to make all the useful interpretations of the animal's movements from the home-range data. Therefore, unless there is a clearly appropriate parametric model, such as the bivariate normal model, an analysis of the data by kernel methods alone will usually suffice.

The assumption of independence made throughout this paper may be tested for using the Swihart and Slade (1985) test. It is suggested that kernel methods could also be applied to data for exploratory purposes in a similar way to the grid cell methods discussed by Voigt and Tinline (1980). This would then avoid problems associated with placing a grid on the study site.

Although kernel methods have been used generally for nonparametric density estimation, in the context of home-range data analysis their use has, unfortunately, been neglected. A possible reason for this is that much of the kernel methods' literature deals with theoretical properties and is quite mathematical. It is hoped that this paper has illustrated the practical importance and potential of kernel methods in home-range data analysis and will encourage their use.

ACKNOWLEDGMENTS

I would like to thank an anonymous referee for helpful comments on an earlier draft of this manuscript.

LITERATURE CITED

Anderson, D. J. 1982. The home range: a new nonparametric estimation technique. Ecology **63**:103–112.

Bowman, A. W. 1985. A comparative study of some kernel-based nonparametric density estimators. Journal of Statistical Computation and Simulation **21**:313–327.

Breiman, L., W. Meisel, and E. Purcell. 1977. Variable kernel estimates of multivariate densities. Technometrics **19**: 135–144.

Cacoullos, T. 1966. Estimation of a multivariate density. Annals of the Institute of Statistical Mathematics **18**:179–189.

Calhoun, J. B., and J. U. Casby. 1958. Calculation of home range and density of small mammals. Public Health Monograph Number 55. United States Department of Health, Education and Welfare, Washington, D.C., USA.

Dixon, K. R., and J. A. Chapman. 1980. Harmonic mean measure of animal activity areas. Ecology **61**:1040–1044.

Don, B. A. C., and K. Rennolls. 1983. A home range model incorporating biological attraction points. Journal of Animal Ecology **52**:69–81.

Epanechnikov, V. A. 1969. Non-parametric estimation of a multivariate probability density. Theory of Probability and Its Applications **14**:153–158.

Ford, R. G., and D. W. Krumme. 1979. The analysis of space use patterns. Journal of Theoretical Biology **76**:125–155.

Jennrich, R. I., and F. B. Turner. 1969. Measurement of non-circular home range. Journal of Theoretical Biology **22**:227–237.

Macdonald, D. W., F. G. Ball, and N. G. Hough. 1980. The evaluation of home range size and configuration using radio tracking data. Pages 405–424 *in* C. J. Amlaner and D. W. Macdonald, editors. A handbook on biotelemetry and radio tracking. Pergamon, Oxford, England.

Parzen, E. 1962. On estimation of a probability density function and mode. The Annals of Mathematical Statistics **33**:1065–1076.

Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics **27**:832–837.

Samuel, M. D., D. J. Pierce, E. O. Garton, L. J. Nelson, and K. R. Dixon. 1985. User's manual for program home range. Forest, Wildlife and Range Experiment Station Technical Report Number 15. University of Idaho, Moscow, Idaho, USA.

Silverman, B. W. 1986. Density estimation for statistics and data analysis. Chapman and Hall, London, England.

Spencer, W. D., and R. H. Barrett. 1984. An evaluation of the harmonic mean measure for defining carnivore activity areas. Acta Zoologica Fennica **171**:255–259.

Swihart, R. K., and N. A. Slade. 1985. Testing for independence of observations in animal movements. Ecology **66**:1176–1184.

Tarter, M., and R. Kronmal. 1970. On multivariate density estimates based on orthogonal expansions. The Annals of Mathematical Statistics **41**:718–722.

Van Winkle, W. 1975. Comparison of several probabilistic home-range models. Journal of Wildlife Management **39**: 118–123.

Voigt, D. R., and R. R. Tinline. 1980. Strategies for analyzing radio tracking data. Pages 387–404 *in* C. J. Amlaner and D. W. Macdonald, editors. A handbook on biotelemetry and radio tracking. Pergamon, Oxford, England.

Worton, B. J. 1987. A review of models of home range for animal movement. Ecological Modelling **38**:277–298.