International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-2, December 2019

An Adaptation of Kernel Density Estimation for Population Abundance using Line Transect Sampling When the Shoulder Condition is Violated

Baker Albadareen, Noriszura Ismail

Abstract*: Kernel estimation is a commonly used method to estimate the population density in line transect sampling. In general, the classical kernel estimator of $f_X(0)$, which is the probability density function at perpendicular distance $x=0$, inclines to be underestimated. In this study, a power transformation of perpendicular distance is proposed for the kernel estimator when the shoulder condition is violated. The mathematical properties of the proposed estimator are derived. A simulation study is also carried out for comparing the proposed estimator with the classical kernel estimators.

***Keywords* : line transect, power-transformation, kernel estimator, shoulder condition.**

I. INTRODUCTION

Line transect technique is an easy approach that is used in practice to estimate population abundance, D, in line transect sampling. The study area under the line transect sampling is distributed as non-overlapping strips of total length L, in which the transect line is randomly placed on the strip. In each transect line, an observer follows the line to detect and count objects, and also to assign perpendicular distances (x) between the transect line and each detected object. An advantage of the method is that it is adequate to record only the perpendicular distances of the detected objects. At the end, the recorded distances $x_1, x_2, \dots x_n$ are used for estimating the probability density function of the random variable X at $x=0$. It should be noted that the population abundance parameter, $D = nf(0)/2L$, can be estimated using $\hat{D} = n\hat{f}(0)/2L$ [1].

Let $g(x)$ be a non-increasing conditional probability function of detecting an object given that it is placed at a perpendicular distance x.

Revised Manuscript Received on December 30, 2019.

***** Correspondence Author

Baker Albadareen*, School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia.

Noriszura Ismail, School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia.

Given that the random sample distances are $x_1, x_2, \dots x_n$, the relation between the probability density function $f(x)$ and the detection function $g(x)$ is $f(x) = g(x)/\int g(u)du$, which indicates that the functions $f(x)$ and $g(x)$ have the same shape [1].In general, the shape of the detection function is related to the distribution shape at $x=0$, and can be categorized into two groups; the distribution that has a shoulder at $x=0$ (meaning that the detection objects are closer to the transect line), and the distribution that do not have a shoulder at $x=0$ (see [1], [2]). Mathematically, the shoulder condition is equivalent to $f'(0)=0$. Numerically, several approaches can be used to test the data whether it belongs to the first or the second category (see [3]). However, in real data set, several studies show that the shoulder condition is violated for the line transect data of specific wildlife society (see [4], [5]). Many approaches were proposed to estimate $f(0)$ in the literature. The most common non-parametric approach is the kernel method which allows the data to make an inference about themselves regardless of the distribution shape. An extended description of the Rosenblatt–Parzen kernel estimation method is originated by [6]. The classical kernel estimator is given by:

$$\hat{f}_X(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), -\infty < x < \infty \quad (1)$$

where h is the bandwidth, and $K(.)$ is the kernel function.

Let $x_1, x_2, \dots, x_n$ be a random sample of non-negative perpendicular distances. Assuming that the kernel function is symmetric, the usual reflection of kernel estimator of $f_X(0)$ for these data is [7]:

$$\hat{f}_X(0) = \frac{2}{nh}\sum_{i=1}^{n} K\left(\frac{x_i}{h}\right) \quad (2)$$

The bias and variance of estimator (2) are:

$$Bias\left[\hat{f}_X(0)\right]=2f_X^{'}(0)h\int_0^{\infty}uK(u)du + O\left(h^2\right) \quad (3)$$

$$Var\left[\hat{f}_X(0)\right]=\frac{4}{nh}f_X(0)\int_0^{\infty}K^2(u)duf + o\left(\frac{1}{nh}\right) \quad (4)$$

By ignoring the small terms $o(.)$ and $O(.)$; the asymptotic mean squared error (AMSE) of $\hat{f}_X(0)$ is:

$$AMSE\left[\hat{f}_X(0)\right]=\frac{4}{nh}f_X(0)\int_0^{\infty}K^2(u)du + \left(2f_X^{'}(0)h\int_0^{\infty}uK(u)du\right)^2 \quad (5)$$

The kernel density estimator is often used in past studies to make a good estimate of $f_X(0)$. As examples, [6] provided a full description of the method, [7] suggested the bias and variance of the estimator in equation (2), [8] proposed several bias reduction techniques for $f_X(0)$, [9] suggested a transformation method of boundary correction in the estimation of kernel density, [10] suggested a semi-parametric transformation kernel density estimator, and [11] derived a new kernel estimator of $f_X(0)$ when the shoulder condition is not valid. Recently, [12] proposed several kernel estimators for $f_X(0)$, and [13] proposed a generalized form for the kernel function which is adaptive to the population density estimation.The objective of this study is to propose a power transformation of perpendicular distance for kernel estimator when the shoulder condition is violated. The mathematical properties of the proposed estimator are derived. A simulation study is also carried out for comparing the proposed estimator with the classical kernel estimators. The reminder of this paper is organized as follows. In section II we propose a new kernel estimator based on power transformation that can be applied when the shoulder condition is violated. In Section III we carry out simulation study for testing and comparing the proposed estimator with the kernel estimator. Finally, we conclude in Section IV.

II. METHODOLOGY

The classical kernel estimator in line transect sampling as shown in equation (2) provides underestimated values in some cases, and produces estimates with large negative bias (see [14]). In this paper, we propose a new kernel estimator based on power transformation that can be applied when the shoulder condition is violated. It should be noted that the kernel estimation based on transformed data were proposed in several studies. Examples of kernel density based on transformation can be found in [15]–[17].

We propose the transformation: $Y=e^{X/w}-1$, where $0\le X\le w$. We apply this non-decreasing transformation function to our data. Let $f_X(x)$ be the original density function and $f_Y(y)$ be the transformed one. The original density value of $f_X(0)$ is obtained through back-transformation of $f_Y(0)$ such that

$$f_Y(y)=f_X\left(e^y-1\right)\left|\frac{dx}{dy}\right| = f_X\left(w\ln(y+1)\right)\frac{w}{y+1},\ w,\ y\ge 0 \quad (6)$$

so that $f_Y(0)=wf_X(0)$ when $x=0,\ y=0$.

The estimation of $f_X(0)$ requires $f_Y(0)/w$ to be substituted with $\hat{f}_Y(0)/w$. Using the kernel estimator shown in equation (2), the transformed kernel estimator $\hat{f}_Y(0)$ is:

$$\hat{f}_Y(0)=\frac{2}{nh}\sum_{i=1}^{n}K\left(\frac{y_i}{h}\right),\ y_i=e^{x_i/w}-1 \quad (7)$$

If the Gaussian kernel function $K(u)$ is used, the values obtained from estimator (2) and estimator (7) converge to zero for $x>w$, considering that w is sufficiently large (such as $w\ge \max(x_i)+4h$). It can be seen that when $\left|x\,\mathrm{m}x_i\right|>4h$, the corresponding value of $K\left(\frac{x\,\mathrm{m}x_i}{h}\right)$ is vanishing.

Assuming $K(u)$ is a symmetric function, the bias and variance of $\hat{f}_Y(0)$ are:

$$Bias\left[\hat{f}_Y(0)\right]=2hf_Y^{'}(0)\int_0^{\infty}uK(u)du + h^2 f_Y^{''}(0)\int_0^{\infty}u^2K(u)du+o\left(h^2\right) \quad (8)$$

$$=2h\left(-wf_X(0)+w^2f_X^{'}(0)\right)\int_0^{\infty}uK(u)du + O\left(h^2\right) \quad (9)$$

Retrieval Number: B6582129219/2019©BEIESP
DOI: 10.35940/ijitee.B6582.129219
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$$Var\left[\hat{f}_Y(0)\right]=\frac{4}{nh}f_Y(0)\int_0^{\infty}K^2(u)\,du+o\left(\frac{1}{nh}\right) \tag{10}$$

$$=\frac{4}{nh}wf_X(0)\int_0^{\infty}K^2(u)\,du+o\left(\frac{1}{nh}\right) \tag{11}$$

If the small terms $o(.)$ and $O(.)$ are ignored, the asymptotic mean squared error of $\hat{f}_Y(0)$ is:

$$AMSE\left[\hat{f}_Y(0)\right]=\frac{4}{nh}wf_X(0)\int_0^{\infty}K^2(u)\,du+\left(2h\left(-wf_X(0)+w^2f_X^{'}(0)\right)\int_0^{\infty}uK(u)\,du\right)^2 \tag{12}$$

III. SIMULATION STUDY

The $AMSE\left[\hat{f}_Y(0)\right]$ shown in (12) assumes that the sample size is large. We carry out simulation study for comparing and testing the proposed estimator with the kernel estimator using different sample sizes, which are $n=50, 100, \text{ and } 200$. The relative bias (RB) and the relative mean error (RME) respectively are $RB=\left\{E\left[\hat{f}(0)-f(0)\right]\right\}/f(0)$ and $RME=\sqrt{MSE\left[\hat{f}(0)\right]}/f(0)$.

We use random samples from two density families that are commonly suggested for line transect method when the shoulder condition is violated. Four detection functions are considered for each density group to cover more possible cases. Altogether there are 8 detection functions. The two density families are:

- Beta (BE) model [18]
  The detection function is
  $g(x)=(1-x)^{\beta},\ 0\le x\le w,\ \beta\ge 1$, and
  $f(x)=(1+\beta)(1-x)^{\beta},\ 0\le x\le w,\ \beta\ge 1$.
  We use parameter values
  $\beta=3.0, 4.0, 5.0 \text{ and } 6.0$, and truncation point
  $w=1$ for the four models.

- Negative exponential model [19]
  The detection function is
  $g(x)=e^{-\beta x}, \beta>0,\ 0\le x\le w$, and
  $f(x)=\beta e^{-\beta x},\ 0\le x\le w$. We use
  $\beta=1.0, 1.5, 2.0, \text{ and } 2.5$ and $w=3.0$ for the four models.

Bandwidth selection

The choice of bandwidth h is critical in the kernel method. Several approaches were suggested in the literature to find the 'optimum' value. One of the leader method is the one that minimizes $AMSE\left[\hat{f}_Y(0)\right]$ by $\frac{d}{dh}AMSE\left[\hat{f}_Y(0)\right]=0$, which is:

$$h=\left(\frac{wf_X(0)\int_0^{\infty}K^2(u)\,du}{2n\left(-wf_X(0)+w^2f_X^{'}(0)\right)^2\left(\int_0^{\infty}uK(u)\,du\right)^2}\right)^{1/3} \tag{13}$$

The performances of the kernel estimator in equation (2) are approximately the same when the symmetric kernel functions based on the mean squared error are used [20]. Therefore, the estimation results using the common kernel functions (Gaussian, biweight, and Epanechnikov) are approximately the same. Therefore, we consider the Gaussian kernel function and use the following estimators for comparison purposes:

- Estimator 1 (Est1): The classical kernel estimator in equation (2) is applied to the original data without transformation using the bandwidth recommended by [6]: $h=1.06\hat{\sigma}n^{-\frac{1}{5}}$, where $\hat{\sigma}=\sqrt{\sum_{i=1}^{n}x_i^2/n}$.
- Estimator 2 (Est2): The proposed kernel estimator in equation (7) is applied to the transformed data using the bandwidth from equation (13). The kernel function $K(u)$ is assumed to be Gaussian distributed, and the unknown values of $f_X(0)\,and\,f_X^{'}(0)$ are estimated using a suitable reference density function (see [2], [6], [21]–[23]). When the shoulder condition is violated, the suitable reference distribution model for $f_X(x)$ is the negative exponential model. Using maximum likelihood estimators $\left(\hat{f}_X(0)=\frac{1}{\bar{x}}\right)$ and $\left(\hat{f}_X^{'}(0)=\frac{-1}{\bar{x}^2}\right)$, the bandwidth is

$$h=\left(\frac{w\left(\frac{1}{\bar{x}}\right)\left(\frac{1}{4\sqrt{\pi}}\right)}{2n\left(-w\left(\frac{1}{\bar{x}}\right)+w^2\left(\frac{-1}{\bar{x}^2}\right)\right)^2\left(\frac{1}{\sqrt{2\pi}}\right)^2}\right)^{1/3}.$$

$$= \left(\frac{\left(\sqrt{\pi}\right)\left(\bar{x}\right)^3}{4nw\left(\bar{x}+w\right)^2} \right)^{1/3}$$

Table 1 and Table 2 show that the transformed estimator (Est2) has smaller absolute RB and RME than the original kernel estimator (Est1) under each density families. Moreover, the RME of (Est2) decrease as the sample sizes increase. Therefore it is concluded that (Est2) is asymptotically more consistent. The results are illustrated in Figure 1.

Table 1. Simulation results of beta model

		n=50		*n=100*		*n=200*	
	Estimator	RB	RME	RB	RME	RB	RME
β=3	Est1	-0.2487	0.2703	-0.2288	0.2433	-0.1986	0.2082
	Est2	-0.0970	0.2625	-0.0904	0.2148	-0.0656	0.1640
β=4	Est1	-0.2681	0.2867	-0.2430	0.2550	-0.2189	0.2276
	Est2	-0.1030	0.2531	-0.0789	0.2093	-0.0638	0.1689
β=5	Est1	-0.2911	0.3090	-0.2614	0.2725	-0.2292	0.2366
	Est2	-0.1247	0.2652	-0.0973	0.2081	-0.0793	0.1709
β=6	Est1	-0.2953	0.3122	-0.2672	0.2783	-0.2428	0.2498
	Est2	-0.1095	0.2578	-0.0937	0.2094	-0.0750	0.1634

Table 2. Simulation results of negative exponential model

		n=50		*n=100*		*n=200*	
	Estimator	RB	RME	RB	RME	RB	RME
β=1.0	Est1	-0.2823	0.3014	-0.2439	0.2561	-0.2135	0.2218
	Est2	-0.1106	0.2658	-0.0644	0.2068	-0.0386	0.1619
β=1.5	Est1	-0.3390	0.3515	-0.3133	0.3216	-0.2788	0.2845
	Est2	-0.1289	0.2614	-0.1054	0.2099	-0.0916	0.1708
β=2.0	Est1	-0.3631	0.3755	-0.3257	0.3333	-0.2982	0.3029
	Est2	-0.1511	0.2670	-0.1096	0.2007	-0.0883	0.1606
β=2.5	Est1	-0.3669	0.3781	-0.3347	0.3418	-0.3078	0.3126
	Est2	-0.1512	0.2624	-0.1177	0.2049	-0.1046	0.1730

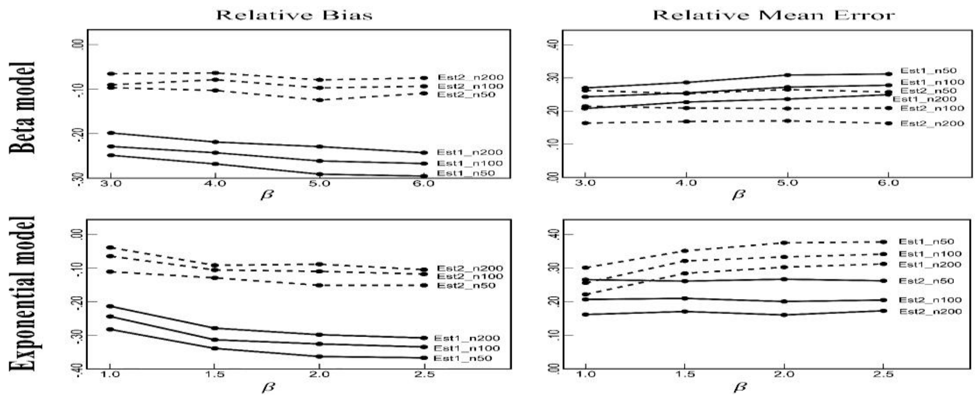

Figure 1. Relative bias and relative mean error of Beta and negative exponential models

IV. CONCLUSION

This study proposed an adaptation of the kernel estimator for population abundance (density) using line transect sampling. The adapted kernel estimator is a power-transformation which is applied to the perpendicular distances. When the shoulder condition is violated, the proposed kernel estimator is shown to be more efficient and consistent than the classical kernel estimator. The asymptotic properties (bias, variance and AMSE) were derived for the proposed estimator. The results of simulation study show that the proposed estimator has superior performance than the classical reflection of kernel estimators, in terms of relative bias, relative mean error for each density family.

ACKNOWLEDGEMENTS

The authors would like to thank the editor and the reviewer for their valuable comments and constructive reviews to further improve this paper.

The authors are also grateful and would like to acknowledge the financial support granted by the Ministry of Higher Education (MOHE) and Universiti Kebangsaan Malaysia (UKM) in the form of research grants (FRGS/1/2019/STG06/UKM/01/5 and GUP-2019-031).

REFERENCES

1. S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas, Introduction to distance sampling: estimating abundance of biological populations, 1st ed. London: Oxford University Press, 2001.
2. Y. P. Mack and P. X. Quang, "Kernel Methods in Line and Point Transect Sampling," Biometrics, vol. 54, no. 2, p. 606, Jun. 1998.
3. S. Zhang, "Generalized likelihood ratio test for the shoulder condition in line transect sampling," Communications in Statistics - Theory and Methods, vol. 30, no. 11, pp. 2343–2354, Jan. 2001.
4. S. T. Buckland, "Perpendicular Distance Models for Line Transect Sampling," Biometrics, vol. 41, no. 1, p. 177, 1985.
5. R. K. Bauer, J. M. Fromentin, H. Demarcq, B. Brisset, and S. Bonhommeau, "Co-occurrence and habitat use of fin whales, striped dolphins and atlantic bluefin tuna in the northwestern mediterranean sea," PLoS ONE, vol. 10, no. 10, pp. 1–21, Oct. 2015.
6. B. W. Silverman, Density estimation: For statistics and data analysis, 1st ed., no. 1. London: Chapman & Hall, 1986.
7. S. X. Chen, "A Kernel Estimate for the Density of a Biological Population by Using Line Transect Sampling," Applied Statistics, vol. 45, no. 2, p. 135, 1996.
8. Y. P. Mack, "Bias-corrected confidence intervals for wildlife abundance estimation," Communications in Statistics - Theory and Methods, vol. 31, no. 7, pp. 1107–1122, 2002.
9. R. J. Karunamuni and T. Alberts, "A locally adaptive transformation method of boundary correction in kernel density estimation," Journal of Statistical Planning and Inference, vol. 136, no. 9, pp. 2936–2960, 2006.
10. G. Koekemoer and J. W. H. Swanepoel, "Transformation Kernel density estimation with applications," Journal of Computational and Graphical Statistics, vol. 17, no. 3, pp. 750–769, 2008.
11. O. M. Eidous, "A new kernel estimator for abundance using line transect sampling without the shoulder condition," Journal of the Korean Statistical Society, vol. 41, no. 2, pp. 267–275, 2012.
12. B. Albadareen and N. Ismail, "Several new kernel estimators for population abundance," AIP Conference Proceedings, vol. 1830, no. 1, p. 80018, 2017.
13. B. Albadareen and N. Ismail, "Adaptive kernel function using line transect sampling," AIP Conference Proceedings, vol. 1940, p. 020112, 2018.
14. O. Eidous, "Variable location kernel method using line transect sampling," Environmetrics, vol. 22, no. 3, pp. 431–440, 2011.
15. L. Devroye and L. Györfi, Nonparametric Density Estimation, 1st ed. New York: John Wiley and Sons, 1985.
16. J. S. Marron and D. Ruppert, "Transformations to Reduce Boundary Bias in Kernel Density Estimation," Journal of the Royal Statistical Society: Series B (Methodological), vol. 56, no. 4, pp. 653–671, 1994.
17. A. Charpentier and E. Flachaire, "Log-Transform Kernel Density Estimation of Income Distribution," L'Actualité économique, vol. 91, no. 1–2, pp. 141–159, 2015.
18. L. L. Eberhardt, "A Preliminary Appraisal of Line Transects," The Journal of Wildlife Management, vol. 32, no. 1, p. 82, Jan. 1968.
19. C. E. Gates, W. H. Marshall, and D. P. Olson, "Line Transect Method of Estimating Grouse Population Densities," Biometrics, vol. 24, no. 1, p. 135, Mar. 1968.
20. M. P. Wand and M. C. Jones, Kernel Smoothing., 1st ed., no. 1. New York: Chapman & Hall, 1995.
21. M. Al-Bassam and O. Eidous, "Combination of parametric and nonparametric estimators for population abundance using line transect sampling," Journal of Information and Optimization Sciences, vol. 39, no. 7, pp. 1449–1462, May 2018.
22. S. X. Chen, "Studying School Size Effects in Line Transect Sampling Using the Kernel Method," Biometrics, vol. 52, no. 4, p. 1283, 1996.
23. S. T. Buckland, "Fitting Density Functions with Polynomials," Applied Statistics, vol. 41, no. 1, p. 63, 1992.

AUTHORS PROFILE

Baker Albadareen is a Ph.D. student of statistics at Universiti Kebangsaan Malaysia (UKM). He obtained his MSc of Statistics in 2011 from Yarmouk University. And received his BSc of Actuarial Sciences in 2006 from the University of Jordan. His areas of interest and research include nonparametric kernel density estimation, abundance estimate using line transect sampling, time series prediction.

Noriszura Ismail is a Professor and Head of Actuarial Science Program in Universiti Kebangsaan Malaysia (UKM). She obtained her BSc and MSc (Actuarial Science) in 1991 and 1993 from University of Iowa, and her PhD (Statistics) in 2007 from UKM. She also passed several papers from Society of Actuaries in 1994. Her current research work focuses on Risk Modelling and Applied Statistics.