ASM Sc. J., 13, 2020
https://doi.org/10.32802/asmscj.2020.sm26(1.18)

Performance of Parametric Model for Line Transect Data

Noryanti Muhammad*, Gamil A.A. Saeed and Wan Nur Syahidah Wan Yusoff

[1]*Faculty of Industrial Sciences & Technology, Universiti Malaysia Pahang, 26300 Gambang, Kuantan, Pahang, Malaysia*

One of the most important sides of life is wildlife. There is growing research interest in monitoring wildlife. Line transect sampling is one of the techniques widely used for estimating the density of objects especially for animals and plants. In this research, a parametric estimator for estimation of the population abundance is developed. A new parametric model for perpendicular distances for detection function is utilised to develop the estimator. In this paper, the performance of the parametric model which was developed using a simulation study is presented. The detection function has non-increasing curve and a perfect probability at zero. Theoretically, the parametric model which has been developed is guar-anteed to satisfy the shoulder condition assumption. A simulation study is presented to validate the present model. Relative mean error (RME) and Relative Bias (RB) are used to compare the estimator with well-known existing estimators. The results of the simulation study are discussed, and the performance of the proposed model shows promising statistical properties which outperformed the existing models.

Keywords: detection function, line transect data, parametric model

I. INTRODUCTION

Distance sampling method has been developed intensively over the last thirty years and has widely applied for estimation of population density (D) especially wildlife populations. In line transect method, a line transect of length L is placed within the interested area, observer travels a randomly placed line and perpendicular distances ($Z_i; i=1,...,n$) are taken from the line to each observed object. The perpendicular distances from the sampled object to line transect represent the sample data set. During the survey course, many objects remain undetected, this can be considered as special characteristic for the line transect method which gives an accurate estimates of population abundance even though not all objects have been detected. The probability of sighted object near the transect line is greater than the probability of sighted object away from the transect line (see also, Buckland *et al.*, 2001).

The paper by Burnham and Anderson (1976) showed that the population abundance D of objects in a specific area satisfies the equation:

$$D = \frac{E(n) f(0)}{2L} \tag{1}$$

where $E(n)$ is the expected value of sighted objects. Burnham and Anderson (1976) introduced the estimated representation of the population density as:

$$\hat{D} = \frac{n\hat{f}(0)}{2L} \tag{2}$$

Equation (2) states that the estimate of $f(0)$ plays the major milestone for estimating the population density of objects.

Let $Z_1,...,Z_n$ be independent-valued random variables of perpendicular distances having common unknown density function $f(z)$ and consider the random sample $Z_1,...,Z_n$ of size n . The conditional function $f(z)$ depend to one of

*Corresponding author's e-mail: noryanti@ump.edu.my

most important concepts in line transect sampling, called detection function which defined as:

$g(z) = P$ (an object is detected / its perpendicular distance is z)

The probability density for a detected distance Z is:

$$f(z) = g(z) \Big/ \int_0^w g(t)dt ; 0 \le Z \le w \tag{3}$$

Equation (3) was pioneered by Burnham and Anderson (1976). One of the ways to deal with line transect technique is that the detection function $g(z)$ should have strictly monotone decreasing for $z \ge 0$. In addition, satisfies the shoulder condition at origin, it can consider mathematically, $g(0) = 1$ and $g'(0) = 0$, it important to refer that the probability density function (pdf) $f(z)$ has the same shape of $g(z)$ and the area under $f(z)$ equal one.

Let $Z_1, \ldots, Z_n$ be a set of perpendicular distances which are usually assumed to be a random sample (Buckland et al., 1993), having a density function $f(z,\theta)$ depends on unknown parameter θ, where θ may one parameter or vector of parameters. Since the $f(0)$ is function of the parameter θ therefore , the estimate of θ lead us to estimate $\hat{f}(0) = f(0,\hat{\theta})$. Gates et al. (1968) presented exponential model, the density function is:

$$f(z,\theta) = \frac{e^{-x/\theta}}{\theta} ; z \ge 0, \theta > 0 \tag{4}$$

The corresponding detection function given as

$$g(z,\theta) = e^{-x/\theta} ; z \ge 0, \theta > 0 \tag{5}$$

The MLE estimate of $f(0)$ is,

$$\hat{f}_{MLE}(0) = \frac{1}{\bar{Z}} \tag{6}$$

where $\bar{Z}$ is the sample mean. It is worth to refer that the detection function $g(z)$ (or the pdf $f(z)$) does not satisfy the shoulder condition. In contrast, the half normal model $f(z)$ achieves the property of shoulder condition. Hemingway (1967) suggested the half normal model with pdf.

$$f(z,\sigma^2) = \sqrt{\frac{2}{\pi\sigma^2}} e^{-z^2/2\sigma^2} ; z \ge 0, \sigma^2 > 0 \tag{7}$$

and the half normal detection function is

$$g(z,\sigma^2) = e^{-z^2/2\sigma^2} ; z \ge 0, \sigma^2 > 0 \tag{8}$$

The MLE is the main estimator to estimate σ^2. Given n perpendicular distances $Z_1, \ldots, Z_n$. The likelihood function based on the half normal model in Equation (7) as term of $f(0)$ can be written as:

$$L(f(0)) = f^n(0) e^{-\frac{\pi f^2(0)}{4} \sum_{i=1}^{n} z_i^2} \tag{9}$$

While, the log likelihood function based on the half normal model in Equation (9) is:

$$\log L(f(0)) = n \log f(0) - \frac{\pi f^2(0)}{4} \sum_{i=1}^{n} z_i^2 \tag{10}$$

To estimate $f(0)$, we maximize the Equation (10) and have:

$$\frac{d \log L(f(0))}{d(f(0))} = \frac{n}{f(0)} - \frac{\pi f(0)}{2} \sum_{i=1}^{n} z_i^2 = 0 \tag{11}$$

Then, the MLE of $f(0)$ is written as

$$\hat{f}_{MLE}(0) = \sqrt{\frac{2}{\pi T}} . \tag{12}$$

Where $T = \frac{1}{n} \sum_{i=1}^{n} z_i^2$. For the half normal model in Equation (7), by using the fact that $T = \frac{1}{n} \sum_{i=1}^{n} z_i^2$ is Gamma distributed. Quinn and Gallucci [15] derived the minimum variance unbiased estimator (MVUE) of $f(0)$ which is given as:

$$\hat{f}_{MVUE}(0) = \frac{1}{\beta(n)} \sqrt{\frac{2}{\pi T}}$$

where

$$\beta(n) = \frac{\Gamma((n-1)/2)}{\Gamma(n/2)} \left(\frac{n}{2} \right)^{1/2} .$$

The shrinkage (SH) estimator is proposed by Zhang (2011) based on half normal model in Equation (7) as:

$$\hat{f}_{SH}(0) = \frac{n-2}{n}\beta(n)\sqrt{\frac{2}{\pi T}}.$$

The estimator $\hat{f}_{SH}(0)$ is biased for $f(0)$, but it achieves the smallest mean square error (MSE).

By comparing the estimators, $\hat{f}_{MVUE}(0)$ and $\hat{f}_{SH}(0)$ we found that:

$$\hat{f}_{SH}(0) = \frac{n-2}{n}\beta(n)\,\hat{f}_{MLE}(0)$$

Based on Magnus et al. (1978), as $n \to \infty$ then $\beta(n) \to 1$ and $\frac{n-2}{n} \to 1$, Therefore, the three estimators asymptotically are equivalent.

There many authors have considered parametric models, for example, Burnham and Anderson (1976), Pollock (1980), Burnham *et. al.,* (1980) and Buckland (1993), Eidous (2004), Al-ababned and Eidous (2012), Al Eibood and Eidous (2017), Saeed *et. al.,* (1986). The related results online transect sampling can be found in Buckland *et. al.,* (2015).

The methodology of this paper can be summarized as: First, we drive the estimate of $f(0)$ based on the proposed model in Equation (15) which is mentioned in Section II (Saeed *et. al.,* 1986).Second, in Section IV, simulation study is done to compare the performance of the proposed estimate $f(0)$ comparing the performance of the proposed estimate $f(0)$ with the existing estimator especially the negative exponential and half normal models which explain in Section IV .We use Relative Mean Error (RME)and Relative Bias (RB) values for the comparing purpose and the smallest values of RME and RB will give better performance. In this simulation study, the perpendicular distances data (Z) have been generated from Hazard-Rate (HR) model.

II. THE MODEL

Let $Z_1,\ldots,Z_n$ be independent and identically distributed (IID) random variables of perpendicular distances and $f(z)$ be the probability density function (pdf) where the $f(z)$ depend on the detection function $g(z)$. The proposed detection function is given by

$$g(z,\sigma^2) = \left(2 - e^{-z_i^2/2\sigma^2}\right)e^{-z_i^2/2\sigma^2};\, z \ge 0, \sigma^2 > 0. \tag{13}$$

and the first derivative of $g(z,\sigma^2)$ (Saeed *et. al.,* 1986) is give as

$$g'(z,\sigma^2) = \frac{2z}{\sigma^2}\left(e^{-z_i^2/2\sigma^2} - 1\right)e^{-z_i^2/2\sigma^2} \tag{14}$$

Since $g'(0,\sigma^2) = 0$, then the detection function $g(z,\sigma^2)$ has a shoulder condition at the origin. In addition, $g(z,\sigma^2)$ satisfies that the probability of sighted object on the line equals one ($g(0,\sigma^2) = 1$). Figure 1 shows the shapes of the detection function for certain values of σ^2. We can easily observe that $\left(e^{-z_i^2/2\sigma^2} - 1\right) < 0$ for all $z \ge 0, \sigma^2 > 0$, then Equation (14) can be shown that $g'(z,\sigma^2) < 0$.

In order to estimate $f(0)$, we determine the corresponding pdf of $g(z,\sigma^2)$. Therefore, the corresponding pdf of $g(z,\sigma^2)$ can be obtained by normalizing the detection function $g(z,\sigma^2)$ as

$$f(z,\sigma^2) = \frac{1}{\mu}g(z,\sigma^2),$$

where $\mu = \int_0^\infty g(z,\sigma^2)dz$ .Hence, the pdf of z , $f(z,\sigma^2)$ (Saeed et al., 1986) is given as

$$f(z,\sigma^2) = \frac{2}{(2\sqrt{2}-1)\sqrt{\pi\sigma^2}}\left(2 - e^{-z_i^2/2\sigma^2}\right)e^{-z_i^2/2\sigma^2} \tag{15}$$

where $z \ge 0$ and $\sigma^2 > 0$ and the $f(0,\sigma^2)$ is

$$f(0,\sigma^2) = \frac{2}{(2\sqrt{2}-1)\sqrt{\pi\sigma^2}} \tag{16}$$

Equation (16) shows that the $f(0,\sigma^2)$ is a function of the parameter σ^2. Therefore, it is enough to estimate σ^2 for estimating $f(0,\sigma^2)$.

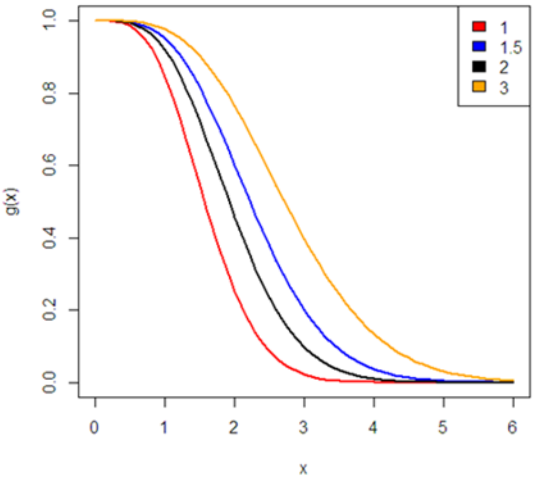

Figure 1. The detection function $g(z)$ of the proposed model for different parameter σ^2

III. MAXIMUM LIKELIHOOD ESTIMATORS

The maximum likelihood estimator MLE for $f(0,\sigma^2)$ can be found by estimating the parameter σ^2 using the MLE. Based on the model $f(z,\sigma^2)$ in Equation (15), the likelihood function $L(\sigma^2)$ given as:

$$L(\sigma^2)=\prod_{i=1}^{n} f(z_i,\sigma^2)$$
$$=\left(\frac{2}{(2\sqrt{2}-1)\sqrt{\pi\sigma^2}}\right)^n \prod_{i=1}^{n}\left(2-e^{-z_i^2/2\sigma^2}\right)e^{-\sum_{i=1}^{n} z_i^2 \big/ 2\sigma^2} \tag{17}$$

and the log likelihood function is

$$\log L(\sigma^2)=\log\left(\frac{2}{(2\sqrt{2}-1)\sqrt{\pi}}\right)^n-\frac{n}{2}\log\sigma^2 -\frac{1}{2\sigma^2}\sum_{i=1}^{n} z_i^2+\sum_{i=1}^{n}\log\left(2-e^{-z_i^2/2\sigma^2}\right) \tag{18}$$

Then, we maximize σ^2 in the Equation (17) as

$$\frac{d\log L(\sigma^2)}{d\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n} z_i^2 -\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}\left(\frac{z_i^2}{2e^{-z_i^2/2\sigma^2}-1}\right)=0 \tag{19}$$

The solution of the Equation (19) leads to obtain the MLE of σ^2, which can be determined by using suitable numerical methods such as Newton-Raphson method.

IV. SIMULATION RESULTS AND DISSECTION

A simulation study is performed to compare the performance between the proposed estimator and the existing estimators which are negative exponential and half normal models. The data of $f(0)$ is simulated from the half normal model which satisfies $f'(0)=0$ and the negative exponential model which does not achieves $f'(0)\neq 0$.The simulation study is based on simulated samples of sizes $n=50,100$ and $n=200$, which consider medium and large sample sizes. For this purpose, we generate the values $Z_1,\ldots Z_n$ which represent the perpendicular distances data from Hazard-Rate (HR) model which is used by Hayes and Buckland (1971). The HR model is given by:

$$f(z)=\frac{1}{\Gamma(1-1/\beta)}\left(1-e^{-z^{-\beta}}\right) \tag{20}$$

In our simulation, four models have been selected from the HR density with parameter values $\beta=1.5,2,2.5,3$ and the corresponding truncation points $\beta=20,12,8,6$,which consider common values used in the literature. The performance of the proposed model is evaluated by using Relative mean error (RME) and Relative Bias (RB) which is defined as:

$$RB=\frac{E\left(\hat{f}(0)\right)-f(0)}{f(0)} \tag{21}$$

and

$$RME=\frac{\sqrt{MSE\left(\hat{f}(0)\right)}}{f(0)} \tag{22}$$

respectively. The values of RB and RME are reported in Table I and Table II for each considered estimator. Notes that the $\hat{f}_{1,MLE}(0)$ is the MLE estimate in Equation (6) which is the exponential model, $\hat{f}_{2,MLE}(0)$ is the MLE estimate in Equation (12) based on the half normal model and the MLE estimate of the proposed model is $\hat{f}_{3,MLE}(0)$ is the MLE estimate of theproposed estimator in Equation (16).

Table 1. RME for the different estimators

n	β	w	$\hat{f}_{1,MLE}(0)$	$\hat{f}_{2,MLE}(0)$	$\hat{f}_{2,MLE}(0)$
50			0.1927	0.5648	0.5154
100	1.5	20	0.1611	0.5685	0.5189
200			0.1523	0.5771	0.5283
50			0.2415	0.4296	0.3663
100	2	12	0.1740	0.4391	0.3733
200			0.1448	0.4382	0.3709
50			0.4636	0.2631	0.2061
100	2.5	8	0.4089	0.2570	0.1801
200			0.3846	0.2593	0.1741
50			0.6121	0.1710	0.1867
100	3	6	0.5781	0.1381	0.1259
200			0.5645	0.1193	0.0881

Table I presents the result of RME for the different estimators when the data are simulated from Hazard-Rate (HR) Model. Based on Table I, we can that the classical estimator, $\hat{f}_{1,MLE}(0)$ is perform the best compared to other estimators $\hat{f}_{2,MLE}(0)$ and $\beta = 1, w = 20$) which the RME gives the smallest values for $\beta = 1, w = 20$ and $\beta = 2.0, w = 12$, regardless of the sample size. The performance of the proposed estimators, $\hat{f}_{3,MLE}(0)$ is outperformed other estimators for $\beta = 2.5, w = 8$ and $\beta = 3.0, w = 6$,regardless of the sample size. Others noticeable finding, the RME values for $\hat{f}_{2,MLE}(0)$) and $\hat{f}_{3,MLE}(0)$ are decreases as the sample size increases. The resulting RME values of this simulation study for the three estimators is presented in Figure 2.

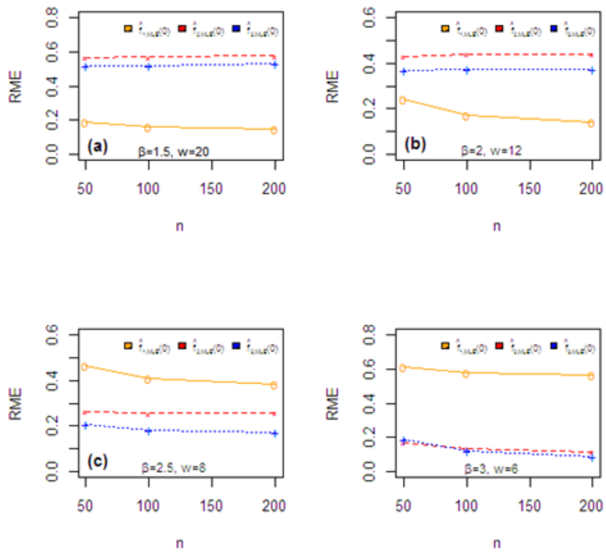

Figure 2. RME values for different estimators

Table II shows the RB values for similar values of β, w and sample size. Based on the table and using the absolute values of RB, the simulation results show similar conclusion as RME where the proposed estimator is outperform other estimators as the sample size and w decreases, and β increases.

Table 2. RB for the different estimators

n	β	w	$\hat{f}_{1,MLE}(0)$	$\hat{f}_{2,MLE}(0)$	$\hat{f}_{2,MLE}(0)$
50			**-0.1077**	**-0.5596**	**-0.5081**
100	1.5	20	**-0.1136**	**-0.5659**	**-0.5153**
200			**-0.1284**	**-0.5759**	**-0.5266**
50			0.1336	**-0.4138**	**-0.3424**
100	2	12	0.1156	**0.4329**	**-0.3639**

200			0.1127	**0.4352**	**-0.3664**
50			0.4114	**-0.2164**	**-0.1178**
100	2.5	8	0.3835	**-0.3713**	**-0.1412**
200			0.3701	**-0.2498**	**-0.1553**
50			0.5702	**-0.0643**	0.0547
100	3	6	0.5585	**-0.0857**	0.0306
200			0.5542	**-0.0926**	0.0230

V. SIMULATION RESULTS AND DISSECTION

In this paper, we have shown that the proposed model in Section II performance by comparing with the existing model. The proposed new parametric model is considered to estimate the population density and it satisfies the property of monotonically decreasing with the detected distances. Furthermore, it achieves the assumption of shoulder condition. Based on the simulation study, the proposed model is very promising to estimate the population abundance using line transect sampling. In addition, the proposed model gives good statistical properties and is recommended to be used to estimate the population abundance.

VI. ACKNOWLEDGEMENT

We gratefully thank to Universiti Malaysia Pahang for financial support under Research University Grant RDU170359.

VII. REFERENCES

Ababneh, F. & Eidous, O. M. 2012, `A weighted exponential detection function model for line transect data', Journal of Modern Applied Statistical Methods 11(1), 144-151.

Al Eibood, F. & Eidous, O. 2017, `A weighted exponential model for grouped line transect data', Mathematics and Statistics 5(1), 1-4.

Buckland, S. (1985), `Perpendicular distance models for line transect sampling' Biometrics pp. 177-195.

Buckland, S., Anderson, D., Burnham, K. & Laake, J. 1993, `Distance sampling: Estimating abundance of biological populations chapman and hall london google scholar'.

Buckland, S. T. 1993, `Fitting density functions with polynomials', Applied Statistics pp. 63-76.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. & Thomas, L. (2001), `Introduction to distance sampling estimating abundance of bio-logical populations.'.

Buckland, S. T., Rexstad, E. A., Marques, T. A. & Oedekoven, C. S. 2015, Distance sampling: methods and applications, Springer.

Burnham, K. & Anderson, D. 1976, `Mathematical models for nonparametric inferences from line transect data', Biometrics pp. 325-336.

Burnham, K. P., Anderson, D. R. & Laake, J. L. 1980, `Estimation of density from line transect sampling of biological populations', Wildlife monographs (72), 3-202.

Gates, C. E., Marshall, W. H. & Olson, D. P. 1968, `Line transect method of estimating grouse population densities', Biometrics pp. 135-145.

Eidous, O. M. 2004. A parametric family for density estimation in line transect sampling. Abhath Al-Yarmouk, 13, 315-326.

Hemingway, P. 1971, `Field trials of the line transect method of sampling large populations of herbivores'.

Magnus, W., Oberhettinger, F., Soni, R. P. & Wigner, E. P. 1967, `Formulas and theorems for the special functions of mathematical physics'. Pollock, K. 1978, `A family of density estimators for line-transect sampling', Biometrics pp. 475{478.

Quinn, T. J. & Gallucci, V. F. 1980, `Parametric models for line-transect estimators of abundance', Ecology 61(2), 293-302.

Saeed, G. A. A., Muhammad, N., Liang, C. Z., Yusoff, W. N. S. W. and Salleh, M. Z. 2017, Model for estimating of population abundance using line transect sampling, in `Journal of Physics: Conference Series', Vol. 890, IOP Publishing, p. 012153. 85

Seber, G. A. 1986, `A review of estimating animal abundance', Biometrics pp. 267{292.

Thomas, L., Buckland, S. T., Burnham, K. P., Anderson, D. R., Laake, J. L., Borchers, D. L. & Strindberg 2002, `Distance sampling', Encyclopedia of environmetrics.

Zhang, S. 2011, On parametric estimation of population abundance for line transect sampling. Environmental and Ecological Statistics, 18(1):79–92.